ADA132048

AFWAL-TR-83-3005

A COMPARISON OF MANUAL AND VOCAL RESPONSE MODES FOR THE CONTROL OF AIRCRAFT SUBSYSTEMS

Anthony J. Aretz, Lt, USAF
Crew Systems Development Branch
Flight Control Division

March 1983

Final Report for Period 1 January 1982 - 30 November 1982

Approved for Public Release; Distribution Unlimited

FLIGHT DYNAMICS LABORATORY
AIR FORCE WRIGHT AERONAUTICAL LABORATORIES
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

DTIC
ELECTE
SEP 01 1983
D
E

Best Available Copy

DTIC FILE COPY

83 08 26 080

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture use, or sell any patented invention that may in any way be related thereto.

This report has been reviewed by the Office of Public Affairs (ASD/PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

ANTHONY J. ARETZ, LT, USAF
Crew Systems Behavioral Engineer
Crew Systems Development Branch
Flight Control Division

TERRY J. EMERSON
Acting Chief
Crew Systems Development Branch
Flight Control Division

FOR THE COMMANDER

MORRIS A. OSTGAARD
Assistant for Research & Technology
Flight Control Division

"If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify AFWAL/FIGR, W-PAFB, OH 45433 to help us maintain a current mailing list".

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFWAL-TR-83-3005	2. GOVT ACCESSION NO.: A132048
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): A COMPARISON OF MANUAL AND VOCAL RESPONSE MODES FOR THE CONTROL OF AIRCRAFT SUBSYSTEMS	5. TYPE OF REPORT & PERIOD COVERED: Final Report 1 Jan 82 - 30 Nov 82
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Lt Anthony J. Aretz	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Flight Dynamics Laboratory (AFWAL/FIGR) Air Force Wright Aeronautical Laboratory Wright-Patterson Air Force Base, Ohio 45433	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 24030447
11. CONTROLLING OFFICE NAME AND ADDRESS: Flight Dynamics Laboratory (AFWAL/FIGR) Air Force Wright Aeronautical Laboratory Wright-Patterson Air Force Base, Ohio 45433	12. REPORT DATE: March, 1983
	13. NUMBER OF PAGES: 120
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Human Factors
Speech Recognition
Multifunction Controls
Time Sharing
Aircraft Controls
Cockpit Controls
Vocal Response
Manual Response
Dual Task Tracking

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The objective of this study was to determine how a vocal response mode compared to a manual response mode for data entry in a fighter cockpit simulator. Specifically, both vocal and manual response modes were compared in single and dual task conditions on the basis of pilot flight performance, response time, and errors while accomplishing several communication, navigation, and weapons tasks. The results indicated that the manual response mode was more effective than the vocal response mode in terms of response time data; however, the vocal response mode was more effective in terms of flying performance data. These results

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 continued.

pointed to a trade off strategy used by the pilots as a function of their current workload. In the dual task manual condition the pilots concentrated on the manual data entry task and flying performance suffered, whereas, in the dual task vocal condition, the pilots kept their attention on the flying task and entered data while maintaining good flight control. As a result, in future fighter aircraft both manual and vocal control should be provided to the pilot for his selection. This conclusion was also supported by questionnaire results in which the pilots favored the implementation of both manual and vocal control.

FOREWORD

This technical report is the result of research performed by the Display Information Interface Group of the Crew Systems Development Branch (FIGR), Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio. Mr. Robert Bondurant was the group leader and Lt Anthony Aretz was responsible for the research. Software support was provided by Mr. Tim Barry and Mr. Steve Haupt of the SCT Corporation. The objective of this effort was to determine how a vocal response compared to a manual response for subsystem control and data entry in a fighter cockpit simulator. This effort was accomplished under Work Unit #24030447.

The final manuscript was typed by Miss Joanne Giorlando of The BDM Corporation, under contract F33615-81-C-3620.

Accession For
NTIS GRA&I X
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Avail and/or
Dist Special
A
DTIC COPY INSPECTED

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

GLOSSARY OF TERMINOLOGY

Electro-Optical Display - A programmable electronic display on which a variety of symbology can be shown.

Kolmogorov Smirnov Test - A one-sample goodness of fit test to determine if the distribution of a set of sample values differs from the normal distribution. Used to analyze questionnaire rating-scale data.

Krishnaiah Finite Intersection Tests (FITs) - A set of tests conducted after significant MANOVA results are found to determine: 1) which of the dependent variables were most sensitive to changes in independent variables; and 2) which of the experimental groups differed significantly from each other.

Multifunction Control (MFC) - Combines several multifunction switches, whose functions change depending upon the task being performed by the operator, on a single panel.

Multifunction Control Logic - The steps by which pilots execute tasks using the MFC.

Multivariate Analysis of Variance (MANOVA) - A statistical procedure which takes into account the fact that several partially correlated dependent variables may be affected by experimental manipulation, and which can determine significant differences in experimental conditions.

Root-Mean-Square Error (RMS) - A summary statistic descriptive of the error amplitude distribution of a sample of tracking performance. Specifically, it is an index of performance variability that is relative to the null point.

SUMMARY

The objective of this study was to determine how a vocal response mode compared to a manual response mode for data entry in a fighter cockpit simulator. Specifically, both vocal and manual response modes were compared in single and dual task conditions on the basis of pilot flight performance, response time, and errors while accomplishing several communication, navigation, and weapons tasks. Subjective responses of pilot preferences were also evaluated. The results indicated that the manual response mode was more effective than the vocal response mode in terms of response time data; however, the vocal response mode was more effective in terms of flying performance data. These results pointed to a trade off strategy used by the pilots as a function of their current workload. In the dual task manual condition the pilots concentrated on the manual data entry task and flying performance suffered, whereas, in the dual task vocal condition, the pilots kept their attention on the flying task and entered data while maintaining good flight control. Therefore, in a low level terrain following phase when flight control of the aircraft is critical, the vocal response mode is probably the most effective alternative since it has the least impact on flight performance. In this situation the pilot would concentrate his information processing resources on the flying task and enter data by voice. In a cruise phase, when flight control is not that critical, the manual response mode is probably more effective since the task can generally be accomplished quicker with fewer errors. As a result, in future fighter aircraft both manual and vocal control should be provided to the pilot for his selection. This conclusion was also supported by the questionnaire results in which the pilots favored the implementation of both manual and vocal control.

SECTION I

Introduction

Rapid advancements in the ability of computers to recognize speech have provided the technology for the application of voice control in the operation of complex systems. For example, planners are now considering the use of voice control in the Air Force's next fighter, the Advanced Tactical Fighter (ATF). The Air Force sees several potential advantages in using voice control in fighter aircraft--the main advantage being the potential for an effective reduction in pilot workload. That is, voice control can provide an alternative response mode for the pilot in controlling the aircraft subsystems, thereby helping to reduce the workload on the already overburdened manual response mode. For instance, in low level flight behind enemy lines pilots are reluctant to take their hands off the stick and throttle and their eyes off visual displays to accomplish manual tasks such as radio frequency changes. In this situation the pilot could use voice control to change the radio frequency and still keep his eyes and hands where they are needed. Lea (1979) lists several other potential advantages of speech recognition in applied settings:

1) Utilizes man's most natural response modality
2) Requires little or no user training
3) Provides for simultaneous communication with machines and other humans
4) Fast, multimodal communication
5) Freedom of movement and orientation
6) No panel space or complex apparatus required
7) Compatible with telephone and radios
8) Eyes and hands are free to perform other tasks

Fighter aircraft are not the only possible application of voice control. In today's rapidly advancing technological world, system operators are often required to

operate systems that place them in a complex multi-task environment that may require the performance of several tasks simultaneously. For example, the operator of an automated assembly line may be required to monitor its status, control its operation, and transmit and receive communications at the same time. If voice control were to be used in a system such as this, the eyes and hands would be freed to accomplish other tasks and overall operator workload would be distributed more efficiently among available response modalities.

Obviously, there are limits on the number of things people can do at the same time. For example, try to listen to one person and talk to another on a different topic at the same time. What is difficult in this situation is that by listening to one person and talking to another, two tasks are competing for the same information processing resources. However, there is evidence that the information processing resources available may not be composed of one resource but several resources--one for visual performance, one for manual performance, one for cognitive performance, etc. (Navon and Gopher, 1979). If this is the case, the advantage of using voice control in applications such as fighter aircraft is that the resources required by voice control come from a different source than the resources required for manual control. If voice control were to be used for tasks that usually require manual and visual resources, then operator workload could be distributed more efficiently among available response modalities.

If voice is going to be used as an alternative response mode, it first must be established that a vocal response is more effective, or as equally effective, as using a manual response for the specific tasks in which the vocal response is to be used. The objective of this study is to determine how a vocal response compares to a manual response in a fighter cockpit environment. Specifically, both vocal and manual response modes will be compared on the basis of pilot

flight performance, response time, and errors while accomplishing several communication, navigation, and weapons tasks in a fighter cockpit simulator. Subjective questionnaire data will also be evaluated.

SECTION II

Literature Review

There are three main topics that need to be discussed to put the current experiment in perspective. First, the operation and performance characteristics of speech recognizers will be discussed. Second, models of human information processing will be reviewed with an emphasis upon how humans use their resources to process information. Finally, prior research that compared manual and vocal response modes will be addressed. By reviewing these three topic areas an overall view of the issues involved in the application of speech recognition in complex systems will be achieved.

Speech Recognition By Computer

Speech recognizers fall into two main categories: discrete word recognizers and continuous speech recognizers. Discrete word recognizers recognize individual words that are separated by a pause. The pause is required in order for the recognizer to tell when one word ends and the next word begins. A discrete word recognizer was utilized in this experiment. Continuous speech recognizers, on the other hand, recognize continuous speech; the requirement that words are separated by a pause is not necessary. The next two sections describe in more detail how these two categories of speech recognizers operate.

Discrete word recognizers. Discrete word recognizers use template matching paradigms to recognize words. The general strategy is to compare the characteristics of an incoming word with template reference patterns stored in the recognizer's memory and determine which template matches the incoming word the best. This comparison results in one of three outcomes: either the word is matched to the correct template (a recognition), it is matched to the wrong template (a misrecognition), or the word is rejected as not matching any of the templates (a rejection).

The recognition of a word by a discrete word recognizer involves three main steps--preprocessing, feature extraction, and comparison (Herscher, 1977). Figure 1 shows these three steps broken down into their components.

The purpose of the preprocessor is to shape the output of the microphone to produce an amplitude and time normalized speech spectrum and to analyze this spectrum using a bank of active bandpass filters. In essence, the preprocessor tries to provide an input to the feature extractor that is free of extraneous noise such as background and breath sounds.

The feature extractor takes the output from the preprocessor and measures it for shapes and changes in the spectrum. Combinations and sequences of these measurements are then processed to produce a set of acoustic features. The acoustic features are then time normalized so that each word is the same length. If the speech recognizer is in the training mode, the time normalized acoustic features for each word are then stored in a template memory. (If more than one template is taken for a single word all the templates for that word will be averaged.)

If the recognizer is in the recognition mode, the time normalized acoustic features for the incoming word are compared to all the patterns stored in memory. The stored template matching the incoming input closest, within a certain tolerance, is recognized as the incoming word. If no match is made within the preset tolerance, the incoming word is rejected as not belonging to the stored set of templates. Obviously, each time the speaker or the vocabulary is changed the speech recognizer must be retrained to create a new set of templates.

Recognition accuracy of discrete word recognizers will vary from around 89 to 99 percent depending upon variables such as vocabulary (size and confusability), noise, stress, or anything else affecting the voice signal input into the recognizer (Reddy, 1976). Highest recognition accuracies

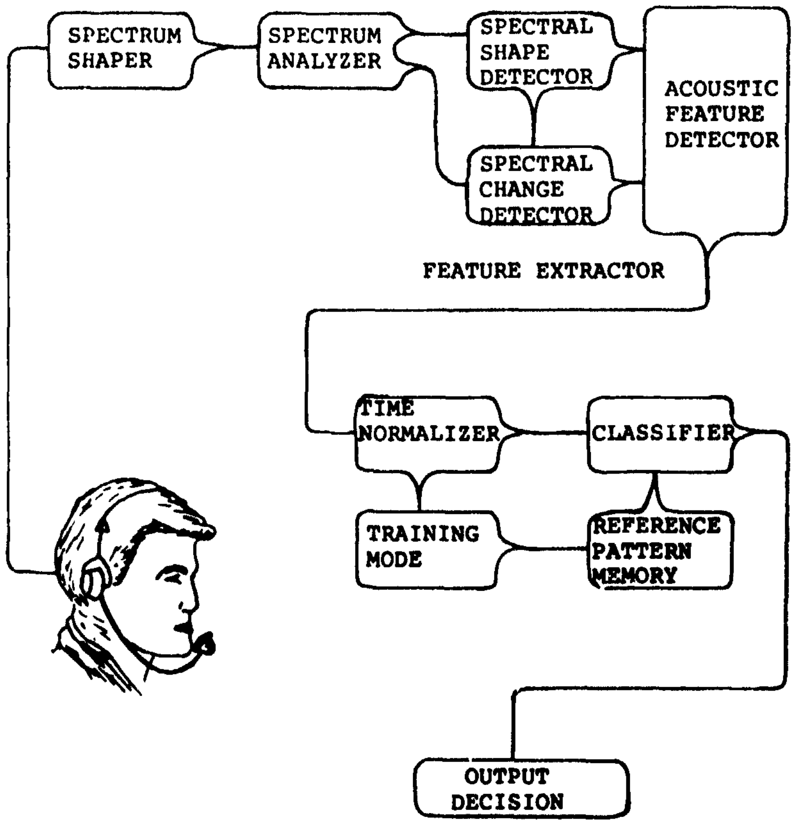

Figure 1. Block diagram of a speech recognizer (From Herscher, 1977).

are obtained in laboratory settings with low noise and robust vocabularies. Doddington and Schalk (1981) did a laboratory study in which they compared the recognition accuracies of currently available discrete word speech recognizers. This study utilized 8 male and 8 female subjects and a 46 word vocabulary (10 words, the digits 1 through 10, and the alphabet from A to Z). The results of the study are shown in Table 1. One interesting point about this data is that performance on all speech recognizers, except for the NEC DP-100, was better for men than for women. This is due to the fact that the male voice, which is lower in frequencies, provides more information about the acoustical features of the vocal response. Doddington and Schalk qualify their data by stating that it is to be used only as a benchmark comparison since the results were obtained in a laboratory with tight experimental control: the acoustic environment was pure and unvarying, the speech level was tightly controlled, and all speech input errors of the speakers were eliminated. In an operational system these results would probably be worse.

Continuous speech recognizers. Continuous speech recognizers are not as accurate in recognizing continuous speech as discrete word recognizers are in recognizing words. Continuous speech recognizers typically achieve recognition accuracies from 55 to 97 percent (Reddy, 1976). The main problem in recognizing continuous speech is the detection of when words begin and end. Reddy describes the state of continuous speech recognition as it existed in 1976:

> We do not yet have good signal-to-symbol transformation techniques nor do we fully understand how to do word matching performance of CSR (Continuous Speech Recognition) systems when compared with word recognition systems. However, researchers have been

Table 1

Evaluation of Speech Recognizers

Manufacturer	Model	Number of Misrecognitions	Errors for Men	Errors for Women
Verbex	1800	10 (0.2%)	2	8
Nippon Electric	DP-100	60 (1.2%)	1.4%	1.0%
Threshold Technology	T-500	73 (1.4%)	1.2%	1.7%
Interstate Electronics	VRM	147 (2.9%)	2.0%	3.7%
Heuristics	7000	300 (5.9%)	4.4%	7.3%
Centigram	MIKE 4725	366 (7.1%)	6.3%	8.0%
Scott Instruments	VET/1	646 (12.6%)	11.2%	14.0%

Note: The data in Table 1 are from Doddington and Schalk, 1981.

working seriously on CSR techniques only for the past few years and significant improvements can be expected in the not too distant future.

Most currently available continuous speech recognizers use a recognition algorithm that processes inputs from left to right. The beginning of the first word is known since no words precede it. When the first match is found, the point at the end of the first word can be assumed to be the beginning of the next word and so on. This technique is not perfect, however, since sounds which link words together can cause recognition errors. For example, the "m" in "some milk" causes the phrase to be recognized as "some" with "ilk" left over. But as Reddy indicated, continuous speech is a relatively new topic area in speech recognition and improvements in recognition accuracy are expected in the near future.

Models Of Human Information Processing

Perhaps the most well known model of human information processing is the "single-channel" model proposed by Broadbent in 1958. This model has been revised by several theorists over the years and even modified by Broadbent himself in 1971. The basic idea of the model is that man has only one information processing channel for which tasks compete for attention. In order for several tasks to be accomplished at the same time, the operator must switch among the tasks and only attend to one task at a time. If the operator tries to accomplish two or more tasks simultaneously, structural interference occurs and the performance on all tasks suffers. In more recent literature, however, the single-channel model has lost support among theorists due to the failure of research to verify the model (Navon and Gopher, 1979). The single channel model is being replaced by a variety of models that provide a basis for the performance of more than one task concurrently.

The more recent models are the result of research which indicates that man can process information in parallel by showing man's ability to attend to more than one task simultaneously. An example of this research is a study by Allport, Atonis, and Reynolds (1972) which showed that skilled piano players could play pieces they had never seen before, while at the same time shadow recorded English prose at a rate of 150 words a minute. Several authors have supported parallel processing models in the literature (Allport, Atonis, and Reynolds, 1972; Eggemeier, 1980; Navon and Gopher, 1979; Norman and Bobrow, 1975; Wickens, Mountford, and Schreiner, 1980). An example of a parallel processing model is one proposed by Kahneman (1973). Kahneman's model includes one central information processing capacity that can be allocated among several ongoing activities concurrently. The processing capacity is allocated among the activities by a mechanism that adjusts the capacity depending upon the state of arousal. Kahneman's model still contains the possibility of structural interference, however, since man still only possesses a limited number of output modalities. The important point about Kahneman's model is that it does provide for the possibility of parallel information processing.

An example of another parallel processing model is described by Navon and Gopher (1979). Navon and Gopher's model is composed of several independent information processing resources in which each resource has a limit on its own capacity as to how much and what type of information it can process. The processing resources can also be allocated among several processes and processing can occur in parallel as well as sequential order. However, Navon and Gopher do not define the resources that are available or how they are allocated among tasks in a time sharing environment.

What parallel information processing models do suggest is that a vocal response , as compared to a manual response, can relax competition for information processing resources in a time sharing environment when the shared task requires manual resources. The reason a vocal response can reduce competition for information processing resources is because a vocal response provides an alternative modality for the output of information other than manual output. In situations where all the information being processed requires manual output, the manual response mode can easily become overburdened. This is the situation in fighter aircraft. However, if some of the information being processed could use a vocal response for output then the burden on the manual response mode can be reduced. This means that operator workload would be distributed more efficiently among the output modalities and overall performance of the system would increase. However, the literature to support this hypothesis is mixed. In a review of the literature on time sharing of verbal and manual tracking tasks, Harris (1978) concludes that

> It is clear that performance of some verbal tasks interferes with simultaneous performance of some tracking tasks. It may be that the requirement to generate a vocal response during tracking is the greatest source of interference. It is not at all clear what other characteristics of verbal tasks may interfere with tracking. Moreover, it is probable that certain parameters of the control task will be important determinants of any decrement in performance observed during simultaneous verbal information processing.

To summarize, it appears that the ability of an operator to do more than one task at a time without a decrement in performance in at least one of the tasks is still open for debate. It seems that no general rules can be derived but it can probably be said that in at least some

situations verbal and manual responses do interfere with each other. Unfortunately, the variables that determine this relationship have not been identified. Therefore, until models of human information processing become more definitive, a comparison must be made for each application of speech recognition to determine if the vocal response interferes with other tasks.

Manual Versus Vocal Response Modes

There has been relatively little research conducted concerning the effectiveness of speech recognition in applied settings. In fact, there were only eight studies found in which manual and vocal response modes were compared on a performance basis. Even though the literature was scarce, the results of these studies generally showed voice input to be an effective response alternative. The results also tended to support parallel processing models of human information processing which would predict better performance when two concurrent tasks were dissimilar in information processing requirements and lower concurrent performance when two tasks were similar in information processing requirements. The purpose of this section is to review some of this research.

All the studies found in the literature were fairly recent publications with the earliest article printed in 1977. This study by Welch (1977) compared a voice input device with two manual entry devices, a typewriter keyboard and a pen and tablet combination. Subjects were required to input two types of data: simple copying of alphanumeric data strings and complex flight data. Speed and accuracy were compared for all three input devices for both types of data. The results showed keyboard entry to be the fastest and most accurate method of entry for alphanumeric data strings. (These results are not surprising since the subjects were highly experienced in using the keyboard.) For the complex data entry task voice was the fastest method of data entry. None of the input methods had significantly

fewer errors. Another variable that Welch investigated was the effect of hand occupation on data entry performance by requiring the subjects to press a button to display the data for input; however, no significant effects were found. Overall, the results of this study showed that voice input is an effective data input modality. Also, the results showed that the complexity of the data to be entered is an important variable to consider.

A study that compared manual vs vocal response modes in a more applied setting was conducted by Taggart and Wolfe (1981). This experiment required subjects to enter preflight data into the stores management or navigation subsystem of a Navy P-3C anti-submarine warfare patrol aircraft. Thirteen subjects made data entries by means of both keyboard and voice control input. The results showed that voice input was significantly faster than keyboard input for the stores management tasks and that manual input was faster than voice input for the navigation tasks. The difference in performance between the two types of tasks was attributed to the fact that the navigation task required character by character input of information, whereas the stores management data was entered using words. Also, some of the subjects in this experiment had prior experience with speech recognition equipment. For these subjects voice input was faster than keyboard input for both types of data entry tasks. The results of this study again point to the effectiveness of voice input as an alternative response mode and that the type of data being entered is an important variable to consider.

Connally (1979) conducted a study in which he compared manual and vocal data entry for entering complete air traffic control (ATC) operational messages. In this study, subjects were required to continuously enter 100 messages typical of the nonradar control position in an ATC center. The messages were written in narrative form on individual cards and required the subject to mentally translate the

messages into a sequence of spoken words for vocal entry or keystrokes for manual entry. This procedure was different from the two previous studies just described which only required the subjects to enter data copied from lists (i.e., requiring no mental translation). The results of this study showed that voice produced fewer errors and saved five minutes on the average mental translation time of the messages as compared to keyboard entry. The savings in mental translation time was attributed to the ability to use natural language for input with a speech recognizer. On the other hand, the keyboard entry had a fifty percent higher data entry rate after mental translation and allowed for easier detection and correction of errors. Taking all the variables together, voice input did not show any clear advantage over manual input of data but could be considered at least as effective as manual input.

In summarizing the research cited to this point, speech input is probably as effective, if not more so, than manual input of data. Also, an important variable to consider in the comparison of manual vs vocal data entry is the type of data being entered. For character by character data, it appears that manual entry will probably be more effective, whereas for complex data involving phrases or words, it appears that vocal entry of data will probably be more effective. Still, these conclusions may not prove to be valid for all applications. The studies described so far only required subjects to enter data. These conclusions may not hold true if subjects were required to perform a manual task, such as tracking, and enter data at the same time. It would appear that in this dual task situation manual data entry would be at a disadvantage since the subjects' hands would be required to perform two tasks simultaneously. The remaining studies in this section describe a few experiments in which subjects were required to perform two tasks at once--a data entry task and a manual tracking task.

Harris, Owens, and North (1979) conducted a study in which they compared single and dual task performance on both a one dimensional compensatory tracking task and a continuous digit-processing task. In the digit processing task the subjects were required to compute the absolute value of the difference between two successive digits and respond with either a vocal or manual input. This task was self paced and the digits were presented either auditorially or visually. The results of this study indicated that both tracking and digit processing performance deteriorated in dual task conditions with no combination of input/output channels equaling the performance in the single task conditions. However, the voice response mode was superior in performance to the manual response mode in the dual task condition when the stimulus was presented visually. When the stimulus was presented auditorially in the dual task condition the manual response mode was more effective, but not as effective as the speech/visual combination. Tracking performance was also significantly better for the vocal response conditions. What these results indicate is that a vocal response can be more effective than a manual response by relaxing the competition for processing resources in a dual task environment when the other task requires manual processing resources. These results also agree with a similar study reported previously by the same authors (Harris, North, and Owens, 1978).

A study by Mountford and North (1980) used a similar design as Harris et al. (1979) except they utilized tasks that were more representative of cockpit data entry. In the experiment, subjects were required to make three self-paced data inputs (selecting a radio, choosing a channel, and entering the data) while accomplishing a continuous compensatory tracking task. Both single and dual task conditions were evaluated. The results showed that in the single task condition manual entry of data was slightly faster than vocal entry; however, in the dual task condition

manual entry time almost doubled, whereas vocal entry time remained the same as single task conditions. Also, tracking performance was poorest in the manual data entry condition. These results somewhat contradict the findings reported by Harris et al. (1979) in showing that performance for the vocal response mode was the same for both single and dual task conditions. However, both studies do support the hypothesis that voice can relax competition for processing resources in a time sharing environment.

Skriver (1979) also compared manual vs vocal entry in single and dual task tracking conditions. Again, the vocal response was superior to manual entry in terms of speed and accuracy. Also, tracking performance was better under dual task conditions for the vocal response mode than for the manual response mode, but dual task tracking performance for either response mode did not equal tracking performance under single task conditions. In addition, Skriver also varied the number of possible response alternatives for data entry. The digits to be entered were presented singly on a CRT display and came from sets with either 4, 8, or 16 response alternatives. The results of the analysis of the data found that the performance for the vocal response, as compared to the manual response, improved as the number of possible response alternatives increased; therefore, the payoff of speech recognition may be dependent on both the type of data being entered and the number of possible entries for any one task.

The effect of varying the difficulty of the tracking task on the performance of the two response modes was investigated in a study by Wickens, Vildulich, and Sandy (1981). In this study the authors varied the level of difficulty of the tracking task (i.e., first vs second order), response mode (visual vs manual), and stimulus (visual vs auditory). The results showed that the tracking task difficulty, as reflected by the control order of the task, influenced the relationship between the effectiveness

of the two response modes. Mainly, a vocal response was better than a manual response when the stimulus was presented visually and the performance for the vocal response mode improved as tracking difficulty increased. Performance for manual data entry was best when the stimulus was presented auditorially but the performance for this combination was still less than the performance for the vocal response for both auditory and visual presentation of the stimulus. Again, these results point to the effectiveness of a vocal response mode when coupled with a manual tracking task in terms of relaxing competition for resources. Also, the results of this study show that the effectiveness of the vocal response mode increases with the level of difficulty of the manual task.

The Problem

Three tentative conclusions can be reached based upon the literature just described. First, the vocal response mode is more effective than the manual response mode for complex data entry tasks. This conclusion is only limited to complex data entry tasks because in simple data entry tasks involving character by character input, the manual response mode appears to be more effective; whereas, for complex data entry involving words and phrases the vocal response mode appears to be more effective. Second, the vocal response mode becomes even more effective than the manual response mode when the data entry task is performed concurrently with another manual task such as tracking. All the research cited that involved a data entry task conducted concurrently with a manual tracking task showed that tracking performance was better when the data was entered using the vocal response mode (Harris, et al., 1979; Mountford and North, 1980; Skriver, 1979; and Wickens, et al., 1981). Third, the vocal response mode is more effective in the dual task environment because the vocal response mode provides a parallel information processing channel that relaxes competition for resources required by

the manual tracking task. All the studies cited involving a tracking task showed that performance for the manual tracking task was better for the vocal response mode condition as compared to the manual response mode under dual task conditions, and that dual task performance for the vocal response mode was not as good as single task tracking performance (Harris et al., 1979; Mountford and North, 1980; Skriver, 1979; and Wickens et al., 1981). These results would suggest that some interference is occurring under dual task conditions for both response modes but that the interference is less for the vocal response mode. Therefore, a vocal response relaxes competition for processing resources required by a manual tracking task, whereas a manual response would increase competition causing performance on both tasks to suffer.

The purpose of this study is to test these three conclusions in the applied environment of an aircraft cockpit simulator. The specific hypotheses of this study are 1) the vocal response mode will be more effective than the manual response mode in accomplishing a variety of aircraft systems tasks (i.e., communication, navigation, and weapons tasks), 2) that the effectiveness of the response modes is related to the complexity of the task, and 3) that speech provides a parallel information processing channel. The third hypothesis that performance for the vocal response mode will be similar in both single and dual task conditions, whereas, performance for the manual response mode will decrease in dual-task conditions.

Task complexity as used in this study refers to the number of switch hits or vocal commands required by any one data entry task. In the research discussed earlier, complexity referred to the type of data being entered (i.e., individual characters vs words). By this convention all of the tasks involved in this study are equally complex since they involve both digits and words. To differentiate

between the tasks used in this study, the number of data inputs required by a task was used to classify the tasks into five different complexity levels described later.

SECTION III

Method

Subjects

Sixteen operationally qualified male Air Force pilots with a mean of 2241 flying hours (the range was from 750 to 5000 hours) served as subjects in this experiment. The pilots also averaged 33.4 years in age, 9.25 years as pilots, and had a variety of experience in flying different Air Force aircraft from cargo to fighters.

Apparatus

Experimental facilities. The overall configuration of the experimental facility is shown in Figure 2. The following paragraphs describe the key components of the facility which impact this study.

Simulator. A single-place cockpit simulator of A-7 geometry containing electro-optical displays (i.e., CRTs) and a Multifunction Control (MFC) was utilized for this evaluation (Figure 3). Three electro-optical displays presented information to the pilot. The head-up display (HUD) presented, in the pilot's forward field of view, flight control information and readouts of the MFC legends corresponding to the selected MFC switches and digits (Figure 4, see Appendix A). Information pertaining to stores onboard the aircraft was presented pictorially on the Stores Status Format (SSF). The various subsystems to be controlled in this study were accessed and monitored either vocally or manually through the MFC. Traditional electro-mechanical round dial instruments presented necessary engine information (e.g., fuel flow) to the pilot. The stick and throttle were located in conventional locations, in the center and on the left console, respectively.

Experimenter's console. The experimenter's console (see Appendix B) provided the experimenter with 1) an array of repeater displays in the cockpit, 2) a display of the

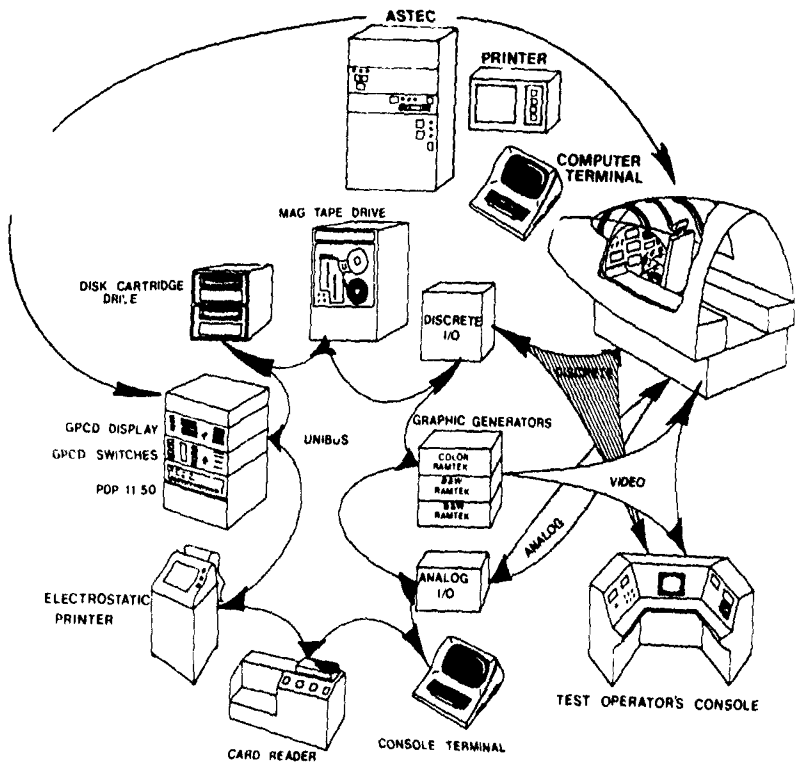

Figure 2. Experimental facility configuration.

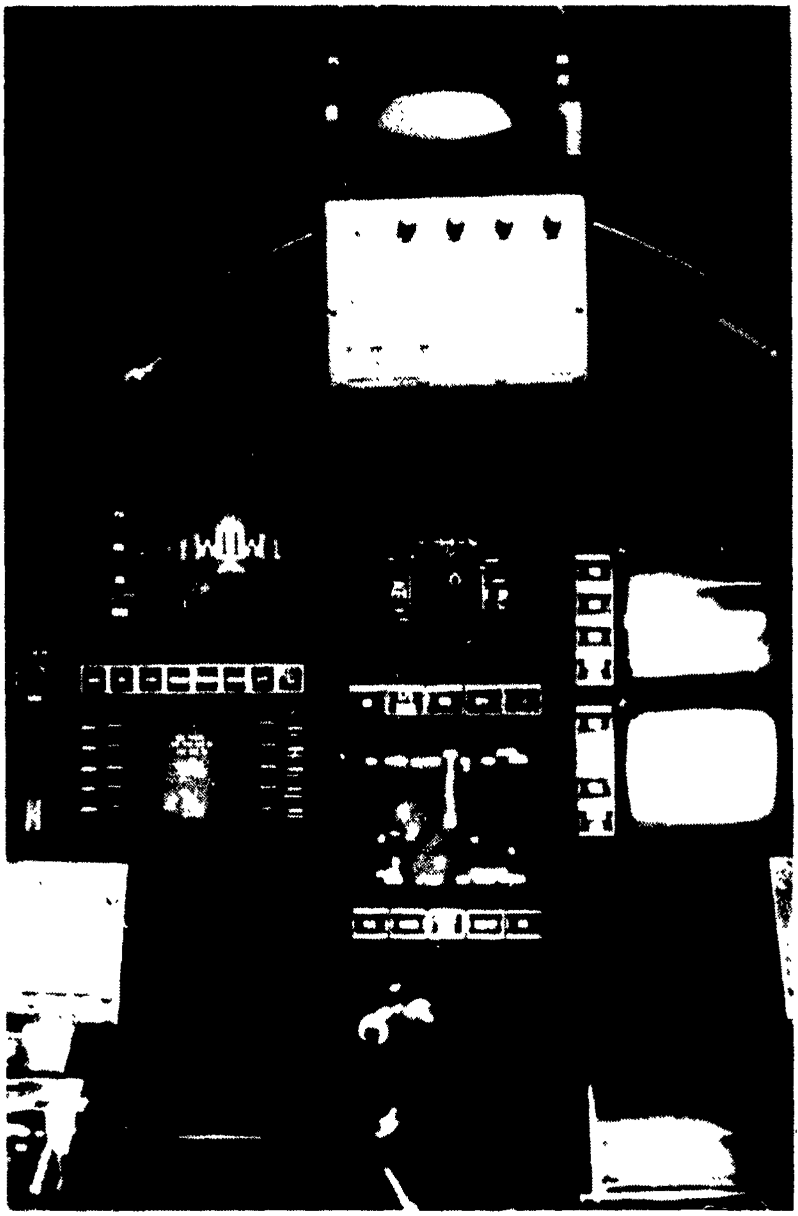

Figure 3. Cockpit simulator.

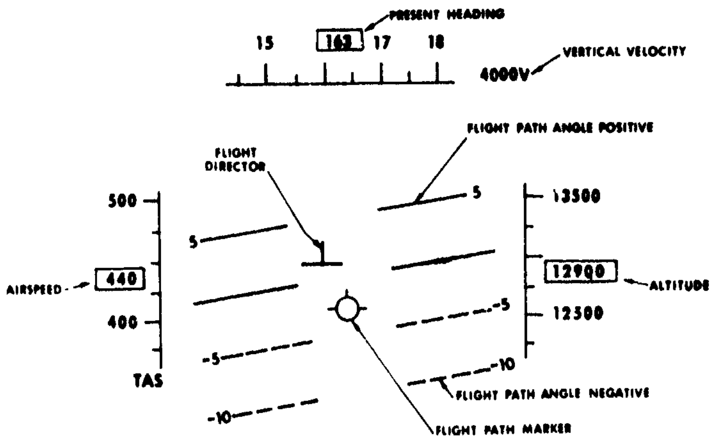

Figure 4. Head-up display.

current experimental status (e.g., task number, response, etc.) and 3) the capability to control the simulation, initiate tasks, and record data.

Multifunction control. The MFC (Figure 5) was located on the left front instrument panel (Figure 3). It consisted of dedicated push button select switches in a row across the top of the CRT and ten multifunction switches mounted in columns on the left and right sides of the CRT. Only seven of the dedicated system select switches had legends displayed on the switch faces and the three left most of these were operable. For the ten multifunction switches, the legends were displayed adjacent to each switch and changed according to the current function the switch was serving. These switches could be activated either manually or vocally. Also, the switches were only operable when a legend was displayed adjacent to a switch and when the experimenter initiated a task. These switches remained operable until the task was completed. The data entry keyboard (DEK), located on the left console, became operable and lighted when the pilot was required to select or enter digits. Once the data was entered, the DEK became inoperable and unlighted. The DEK consisted of twelve dedicated push button keys arranged in a 4x3 telephone keyboard layout with the clear and enter keys located on the left and right sides of the zero (Figure 5). For some tasks, the X and Y could be selected on the 7 and 9 keys, respectively.

In tasks that activated the DEK, a pre-entry readout of each digit selected was displayed to the pilot on the MFC and HUD. When the pilot selected the last digit, the pre-entry readout flashed until the pilot entered the data. The pre-entry readout provided the pilot with the capability to verify that the digits selected were accurate. If the pilot made an error that was in the appropriate range or realistic for the task (for example: 236.7 instead of 236.6 for a UHF frequency), the pre-entry readout indicated the

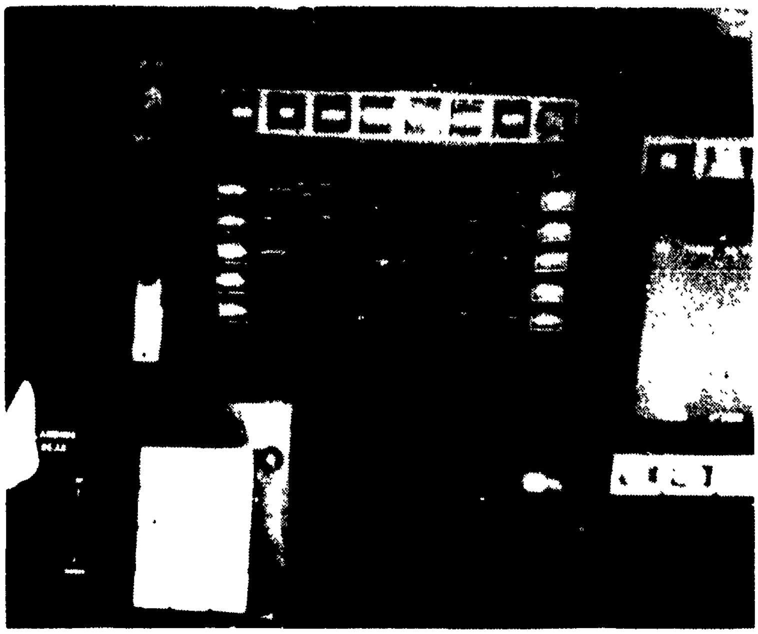

Figure 5. Multifunction control.

incorrect frequency (i.e. = 236.7). In order to correct the mistake, the pilot had to clear the incorrect digit.

If an MFC task was completed incorrectly, the pilot was required to redo the task at the end of the flight. The experimenter was notified of a task error on the experimenter's status display. When an error occurred, the keyboard locked up with the last page of legends used before task completion still displayed to the pilot. After the pilot was notified by the experimenter that an error was made and how to do the task properly, the experimenter continued the flight. The pilot then redid the task after all other tasks for that flight had been completed.

The MFC tasks the pilot was required to perform are shown in Figure 6. Table 2 lists the MFC tasks according to the number of commands required for each task and classifies them into five complexity levels based on the number of commands required for each task. The complexity classification was done in order to investigate the effects of task complexity on the performance of speech recognition since prior research has shown that the type of data being entered influences the performance of speech when compared to manual data entry.

Speech recognizer. The speech recognizer used in this experiment was custom designed by Logicon, Inc., for this facility and is called ASTEC (Advanced Speech Technology Experimental Configuration). Figure 7 shows a block diagram of the seven major components of ASTEC. (The speech digitizer was not used in this study.) The key component of ASTEC is a Threshold 500 speech recognizer that can recognize up to 64 words using the template matching technique described earlier. Recognition accuracies for the Threshold 500 reported in the literature have typically been between 97 and 99 percent (Armstrong and Poock, 1981; Doddington and Schalk, 1981; Poock, 1981; Taggart and Wolfe, 1981).

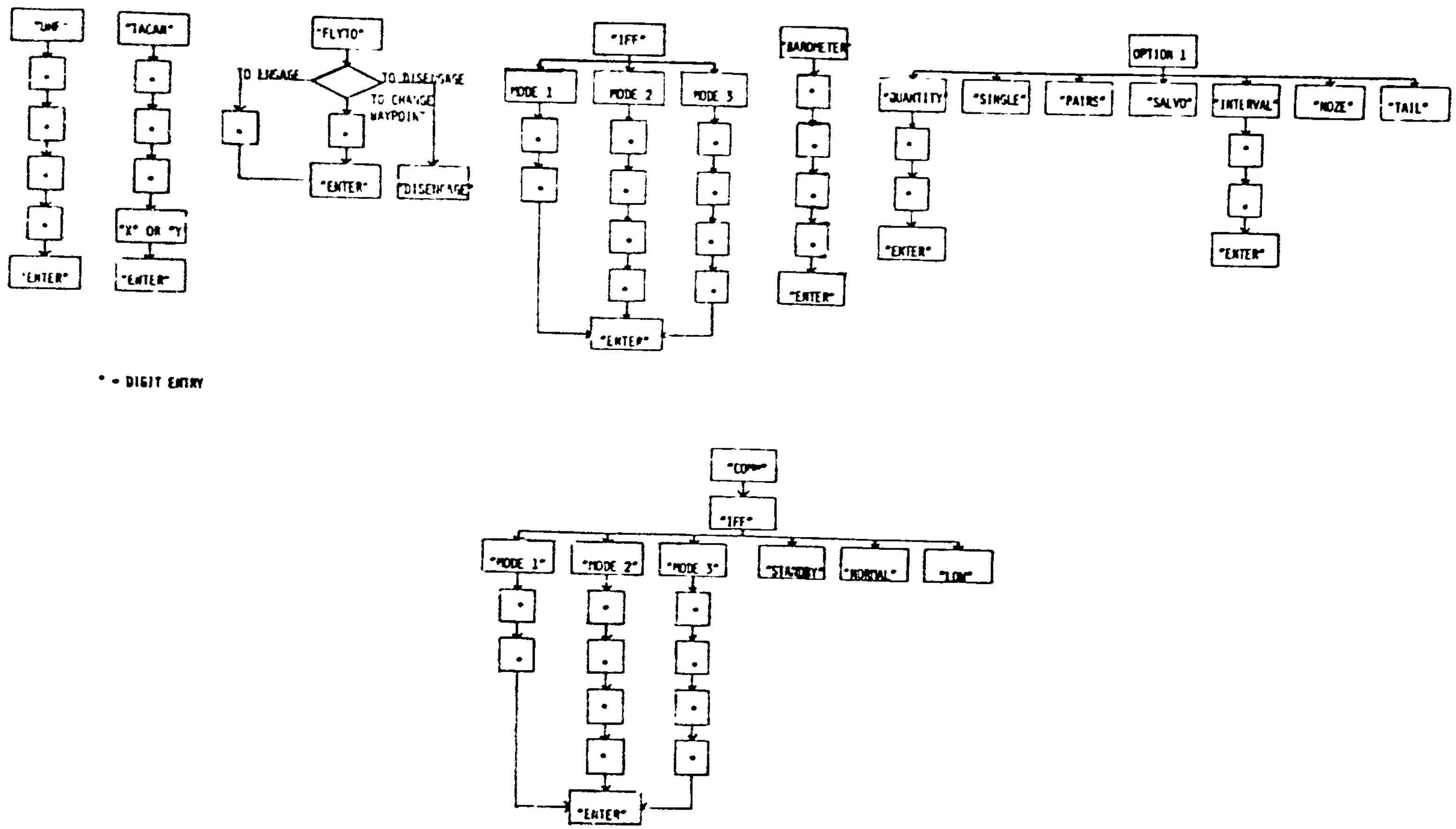

Figure 6. Multifunction control tasks used in this study.

Table 2

Task Complexity Levels

TASK	NUMBER OF COMMANDS	COMPLEXITY LEVEL
Weapon Drop Mode Change	2	1
Weapon Fuzing Change	2	1
Fly to Disengage	2	1
Fly to Engage	3	2
Fly to Change	3	2
IFF Normal	3	2
IFF Mode 1 Change	5	3
Weapon Interval Change	5	3
Weapon Quantity Change	5	3
UHF Change	6	4
TACAN Change	6	4
Barometer Change	6	4
IFF Mode 3 Change	7	5

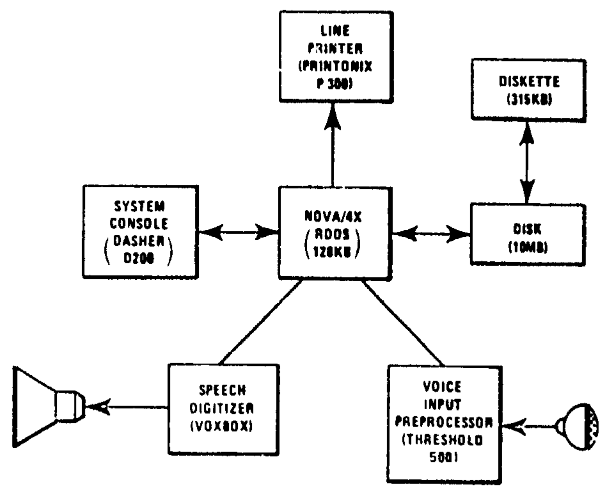

Figure 7. Block diagram of ASTEC.

The configuration of the ASTEC, however, did present one problem - there was an unnecessary lag in the response time that delayed feedback to the pilot when a word was recognized. This response lag was mainly due to the communication between the NOVA computer in ASTEC and the PDP 11/50 computer that controlled the MFC and simulator. This communication took place over an RS-232 port utilizing 10 bit words and a 9600 baud rate. In an operational system the processing that occurred in the NOVA would occur in the main computer and there would be no problem, but for this study there was no room in the PDP 11/50's memory for the word recognition processing that occurred in the NOVA. Even if this study had used a system without this problem, the speech recognizer would still possess an inherent response lag that would not be found in a manual system. This response lag is a result of two processes. First, in order for a speech recognizer to recognize a word it has to detect the end of the word. To do this the speech recognizer looks for a pause at the end of a word. In essence, processing is delayed until the recognizer is sure that the input is a complete word. For the Threshold 500 used in this study, the manufacturer states that this delay is approximately 100 msec. Second, a certain period of time is involved in processing the input to recognize it as a member of the stored vocabulary. The time involved for this process is approximately 150 msec. In addition to the time required by these two processing requirements, the length of the spoken word may also be considered as part of the response time of a speech recognition system.

A study was conducted to determine the actual value of the response lag for the ASTEC system used in this study by measuring the difference between the response times for the MFC when it was operated manually and by speech. The timing study was accomplished by building a circuit which activated a tone at the same time a switch was depressed on the MFC. The recognizer was trained to recognize the tone as a word

corresponding to the activated switch on the MFC. When the PDP 11/50 received the input from the manually activated switch, the computer activated a timer that stopped when a signal was received from ASTEC. The resulting time was the difference between the response times for the two methods of activation, including the length of the tone. However, since the length of the tone was an electronically controlled duration of 1 second, it could be subtracted from the response lag. The difference between the response times for the two methods of activation, after the length of the tone had been subtracted, for 40 observations was found to have an average value of .661 seconds with a standard deviation of .009 seconds. If the 100 msec. required for word detection and the 150 msec. required for word recognition were subtracted from this value, the resulting response lag due to the hardware configuration of the speech recognizer in this study is approximately 400 msec.

Experimental Design

The experimental design used for this study was a 2 x 2 x 5 repeated measures factorial design. The independent variables were: response mode (speech vs manual), task loading (single vs dual task performance), and task complexity level (Table 2). Performance for each pilot was recorded for each condition and the order in which the conditions were presented to the pilots is shown in Appendix C. The treatment order was determined by the use of balanced latin squares so that any treatment was preceded equally often by each of the other treatments. The order of the missions was balanced so that each mission occurred with each treatment an equal number of times. The matrix numbers were a programming tool used to establish the experimental conditions in the computer.

Performance Measures

The pilot's performance for flying the simulator and MFC tasks was recorded by the computer. The following dependent measures were recorded: 1) vertical tracking error

in pixels, 2) horizontal tracking error in pixels, 3) airspeed deviation from the commanded airspeed in knots (i.e. 420 knots), 4) time to initiate an MFC task (time from the completion of the experimenter's instructions to the first input for the task), 5) event time (time from the first input to the last input), 6) total response time (time to initiate plus event time), and 7) MFC errors (number of inputs made minus the number of inputs required for each task).

Procedures

The approximate daily schedule used for each pilot is shown in Table 3. A description of each activity follows in the next several paragraphs.

Simulator briefing. Prior to running the experiment, pilots received a familiarity briefing on the operation of the simulator. The information explained or demonstrated during the briefing included: a) symbology and dynamics of the display formats, b) MFC operation and logic, c) pilot's tasks, d) experimental procedures, and e) flight operation of the simulator.

Speech recognizer training. After the simulator briefing the pilot was trained on the operation of the speech recognizer. This session included training the speech recognizer to recognize the pilot's voice. The training required the pilot to repeat each word to be recognized five times and each digit to be recognized ten times, into the recognizer (Table 4). The digits were trained ten times since their utterances are shorter in duration and are more difficult to recognize.

Training and data missions. Immediately prior to each data flight, a training flight was conducted to give the pilot experience with the handling qualities of the simulator and the procedures for the tasks in the upcoming data flight. The training flights were identical to the data flights except they utilized different missions. The missions used for the training and data flights are shown in

Table 3

Daily Simulation Schedule

TIME	ACTIVITY
0730 - 0800	Simulation Preparation
0800 - 0830	Simulator Familiarity Briefing
0830 - 0900	Train Speech Recognizer
0900 - 0910	Break
0910 - 0930	Training Mission
0930 - 0950	Data Mission
0950 - 1000	Break
1000 - 1020	Training Mission
1040 - 1100	Data Mission
1100 - 1110	Break
1110 - 1130	Training Mission
1130 - 1150	Data Mission
1150 - 1250	Lunch
1250 - 1310	Training Mission
1310 - 1330	Data Mission
1330 - 1430	Data Reduction and Questionnaire Administration

Table 4

Speech Recognizer Vocabulary List

1.	"1"	17.	"Singles"
2.	"2"	18.	"Pairs"
3.	"3"	19.	"Nose"
4.	"4	20.	"Tail"
5.	"5"	21.	"IFF"
6.	"6"	22.	"Mode 1"
7.	"7"	23.	"Mode 3"
8.	"8"	24.	"FLY TO"
9.	"nin-er"	25.	"Normal"
10.	"0"	26.	"TACAN"
11.	"Clear"	27.	"UHF"
12.	"Enter"	28.	"Barometer"
13.	"Option 1"	29.	"Xray"
14.	"Interval"	30.	"Yankee"
15.	"Quantity"	31.	"Disengage"
16.	"Comm"		

Tables 5 and 6 respectively. The order of the tasks was randomized for each mission with one restriction--the "Fly To" tasks had to occur in an established sequence (i.e., fly to engage had to occur prior to fly to change which had to occur prior to fly to disengage). The bomb load used for all missions was identical and is shown in Figure 8. The initial weapon options used for training and data flights are shown in Table 7.

Pilot's tasks. In the single task condition the pilot was required only to operate the MFC (manually or vocally). In the dual task condition, the pilot was also required to fly the simulator in addition to operating the MFC. In flying the simulator, the pilot's task was to keep the velocity vector symbol centered around the flight director symbol on the HUD (Figure 4). The dynamics of the flying task were similar to those of a pursuit tracking task. The pilot flew a programmed terrain following flight path with a constant heading. A gust model was also programmed into the flight director to introduce some random error. This made the tracking task more realistic in simulating actual flight. The pilot was also required to maintain 420 knots airspeed. These two tasks combined were meant to simulate the visual and manual task loading typically experienced during a terrain following segment of a normal fighter aircraft mission.

Single task flying performance was recorded during the dual task condition for a 15 second period prior to the experimenter's instructions for each MFC task.

Simulation flight procedures. The procedures to be described below were identical for both training and data flights except that no data was recorded during the training flights and a different set of missions was used. The procedures for the MFC tasks in the single task condition will be explained followed by a description of how the MFC tasks were accomplished during a flight.

Table 5

Training Missions

TASK NUMBER	I	II	II	IV
1	Fuzing (N)*	FLY TO I (3)	VHF (285.6)	Quantity (14)
2	IFF N/L/S (Normal)	Barometer (29.96)	Barometer (29.96)	TACAN (112Y)
3	TACAN (112Y)	Quantity (14)	FLY TO I (3)	Drop Mode (PAIRS)
4	UHF (285.6)	IFF Mode 1 (42)	Interval (30)	Interval (30)
5	IFF Mode 3 (4041)	FLY TO II (4)	IFF N/L/S (Normal)	FLY TO I (3)
6	FLY TO I (3)	UHF (285.6)	IFF Mode 3 (4041)	Barometer (29.96)
7	Interval (30)	Drop Mode (Pairs)	IFF Mode 1 (42)	Fuzing (N)
8	FLY TO II (4)	IFF N/L/S (Normal)	Drop mode (Pairs)	FLY TO II (4)
9	FLY TO III	FLY TO III	TACAN (112Y)	FLY TO III
10	Drop Mode (PAIRS)	TACAN (112Y)	FLY TO II (4)	IFF Mode 1 (42)
11	IFF Mode 1 (42)	Fuzing (N)	Fuzing (N)	IFF N/L/S (Normal)
12	Barometer (29.96)	IFF Mode 3 (4041)	Quantity (14)	UHF (285.6)
13	Quantity (14)	Interval (30)	FLY TO III	IFF Mode 3 (4041)

* () Indicates entry to be made.

Table 6

Data Missions

TASK NUMBER	I	II	III	IV
1	FLY TO I (3)*	TACAN (115X)	Quantity (12)	Fuzing (T)
2	TACAN (116Y)	IFF N/L/S (Normal)	Barometer (29.82)	Interval (50)
3	FLY TO II (4)	Barometer (30.02)	FLY TO I (3)	IFF Mode 1 (21)
4	Drop Mode (Singles)	UHF (303.4)	IFF N/L/S (Normal)	FLY TO I (3)
5	Interval (75)	Drop Mode (Singles)	Interval (25)	Barometer (29.98)
6	Quantity (12)	Fuzing (T)	FLY TO II (4)	IFF N/L/S (Normal)
7	Fuzing (Nose)	Quantity (18)	IFF Mode 3 (3600)	Quantity (18)
8	IFF Mode 3 (2100)	IFF Mode 3 (6500)	FLY TO III	IFF Mode 3 (4500)
9	FLY TO III	FLY TO I (3)	TACAN (110Y)	FLY TO II (4)
10	Barometer (29.92)	Interval (50)	Fuzing (N)	Drop Mode (Singles)
11	IFF N/L/S (Normal)	FLY TO II (4)	Drop Mode (Singles)	TACAN (101X)
12	IFF Mode 1 (51)	IFF Mode 1 (62)	UHF (283.6)	FLY TO III
13	UHF (348.0)	FLY TO III	IFF Mode 1 (33)	UHF (323.0)

* () Indicates entry to be made.

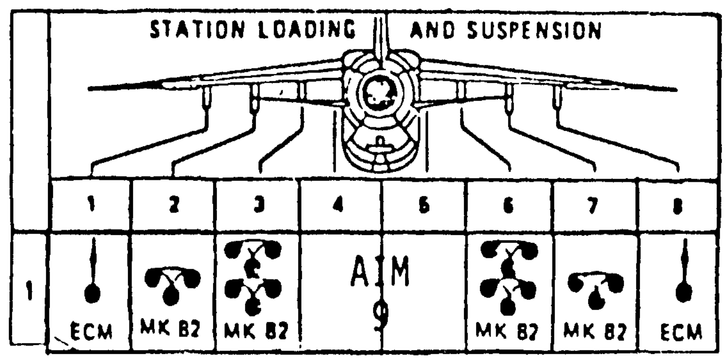

Figure 8. Bomb load used for all missions.

Table 7

Initial Bomb Loads

Training Missions*	Data Missions*
Option 1	Option 1
10 MK-82s	10 MK-82s
No Fuzing	No Fuzing
Singles Drop Mode	Salvo Drop Mode
100 ft. Interval	100 ft. Interval
Master Arm On	Master Arm On
2 ECM Pods On	2 ECM Pods On

* The aircraft is always fully loaded with 18 MK-82s. The options shown are the initial conditions for each mission.

The MFC tasks began with instructions from the experimenter for the task to be accomplished. For example, the experimenter would say: "Boxcar 11 (the pilot's call sign) change your UHF to 236.8." These instructions were read from a script constructed around the mission tasks (see Appendix D). At the completion of this command, the experimenter pushed an EVENT switch on the experimenter's console that started the recording of elapsed time. When the pilot made the first input, either manually or vocally, the timer stopped recording the time to initiate and started recording the event time. At the completion of the last required entry the computer stopped recording elapsed time.

During the dual task condition, when the pilot was flying the simulator, the procedures were slightly different. The difference was that the experimenter pushed a PRE-EVENT switch on the console which started a fifteen second timer; another push of the switch started the timer over again. Activation of this switch initiated data recording of single task flight performance. The criteria the experimenter used to initiate the recording of this data were: the pilot had to be on target and within ±5 knots of the command airspeed. Once the 15 second pre-event period was completed the experimenter initiated the next MFC task which used the same procedures as single task conditions except flight performance data was recorded during the task.

After the pilot completed all the required tasks for a mission successfully, the experimenter terminated the flight by pushing the MISSION COMPLETE switch on the console. After the flight was terminated, a summary statistics program was run to insure that all data had been recorded for each task.

A final debriefing questionnaire was administered following the completion of all data flights and was designed to elicit subjective pilot evaluation of speech recognition as an alternative response mode in the cockpit.

SECTION IV

Results

The data gathered in the study were analyzed in two separate analyses -- one for response time data and one for flying performance data. The results of these analyses are described in the following sections.

Response time data analysis

There were two different measures of response time recorded in this study that are of primary interest in the analysis of the data: 1) time to initiate a task and 2) event time. Time to initiate a task was measured from the end of the experimenter's instructions for the task to the first switch hit (manual) or vocal command of the pilot. Event time was measured from the first switch hit or vocal command to the final switch hit or vocal command required for proper completion of the task. Total response time was also recorded but since this measure was equal to the sum of time to initiate a task and event time it was not included in the analysis. One point to remember is the data analyzed in this section contains the response lag of the speech recognizer described earlier and that this lag accumulates for each data entry.

Time to initiate a task and event time were first analyzed using the General Linear Model (GLM) program containing a Multivariate Analysis of Variance (MANOVA) procedure for unbalanced data. (The unbalanced procedure was used since in using the raw data the number of observations for task complexity level 5 had one third fewer observations than the other task complexity levels.) The GLM program is contained in the Statistical Analysis System (SAS), (Helwig and Council, 1979). A 2 x 2 x 5 repeated measures MANOVA was run on the response time data and the results of this analysis are shown in Table 8. All the main effects (response mode, task loading, and task complexity level) and two interaction effects (response mode x task

Table 8

Multivariate Analysis of Variance
Source Table for Time to Initiate a Task
and Event Time

Source	d.f.	F (Pillai's Trace)	P
Response Mode (Manual vs Speech)	2, 14	33.47	$p < .0001$
Task Loading (Single vs Dual)	2, 14	5.04	$p = .0225$
Task Complexity Level	8, 120	12.55	$p < .0001$
Response Mode x Task Loading	2, 14	6.64	$p = .0094$
Response Mode x Complexity Level	8, 120	5.19	$p < .0001$
Task Loading x Complexity Level	8, 120	0.42	$p = .9055$
Response Mode x Task Loading x Complexity Level	8, 120	1.03	$p = .4156$

loading and response mode x task complexity level) were found to be statistically significant. Table 9 shows the means for the significant main effects and Table 10 shows the means for the significant interaction effects. Several Finite Intersection Tests (FIT), (Cox, Krishnaiah, Lee, Reising, and Schuurman, 1980) were run to determine where the significant differences were located among the means for each significant effect found in the MANOVA. FIT is a simultaneous multiple comparison statistical technique that determines where the significant differences lie among the levels of an independent variable and which dependent variables account for these differences. The results of the FIT analyses are described in the following paragraphs.

Response mode. The first FIT examined the differences among the means for the response mode main effect. The results indicated that both time to initiate a task and event time contributed significantly to the main effect, $F(1,829)=122$, p $.025$ and $F(1,830)=14.7$, p $.025$, respectively. Based on these results it can be concluded that manual data entry was significantly faster than vocal data entry for both time to initiate a task and event time. It should be remembered that these times do include the response lag for the speech recognizer described earlier and a correction for this lag will be examined later on.

Task loading. The next FIT examined the differences among the means for the task loading main effect. The results of this analysis showed that only mean time to initiate a task was significantly different between the single and dual task conditions, $F(1,829)=20.3$, p $.025$. Mean event time was not significant, $F(1,830) = 4.91$.

Task complexity level. The FIT that analyzed the task complexity level main effect found that all the levels significantly differed from each other except for levels 3 and 4. Event time was found to be the only dependent variable that significantly contributed to the differences among the levels. Time to initiate a task was not

Table 9

Means for Time To Initiate a Task and Event Time

Condition	Time To Initiate A Task	Event Time
Response Mode		
Manual	2.48	5.21
Speech	4.06	6.29
Task Loading		
Single	2.92	5.44
Dual	3.62	6.07
Task Complexity Level		
1	3.22	2.75
2	3.10	4.64
3	3.23	7.38
4	3.34	7.21
5	3.80	8.97

NOTE: Times are in seconds.

Table 10

Means for Time to Initiate a Task and Event Time for Interaction Effects

Condition	Time To Initiate A Task		Event Time	
	Speech	Manual	Speech	Manual
Task Loading				
Single	3.99	1.84	6.21	4.67
Dual	4.13	3.11	6.38	5.76
Task Complexity Level				
1	4.07	2.36	2.46	3.05
2	3.72	2.48	4.51	4.76
3	4.14	2.32	8.06	6.69
4	4.14	2.56	8.61	5.85
5	4.51	3.08	11.19	6.74

Note: Times are in seconds.

significantly different among the levels. Table 11 shows the results of this analysis and Figure 9 shows a plot of mean time to initiate a task and mean event time by task complexity level. As Figure 9 shows, mean event time generally increased as task complexity level increased, whereas, mean time to initiate a task remained relatively constant. The overall increase in mean event time would be expected since each increase in the level of task complexity required more inputs to complete a task.

Response mode by task loading. The results of the FIT that examined the significant interaction between response mode and task loading were a little more complex. For time to initiate a task, both single and dual task performance for the manual response mode were significantly faster than the corresponding performance for the single and dual task vocal response mode. Mean time to initiate a task for the manual response mode was also significantly faster for the single task condition than for the dual task condition. These results are shown in Table 12 and are depicted in Figure 10. For task event time, the only significant difference found was between single task performance for the manual and vocal response modes with the manual response mode being faster of the two, $F(1,828)=15$, $p < .025$. Performance for the vocal response mode was not significantly different for single and dual task conditions. Figure 11 shows a plot of these results. One point of interest in Figure 11 is that the difference between dual task performance for the two response modes was not significantly different, indicating that although manual input was faster under single task conditions, it was not significantly faster under dual task conditions.

Response mode by task complexity level. The FIT that examined the significant interaction between response mode and task complexity level showed that for all the levels of task complexity, except level 5, mean time to initiate a task was significantly faster for the manual response mode

Table 11

Results of the FIT Analysis on Task Complexity Level

Comparison Between Levels	d.f.	F	P
1 vs 2	1,827	27.107	$p < .025$
2 vs 3	1,827	57.385	$p < .025$
3 vs 4	1,827	.196	n.s.
4 vs 5	1,827	11.662	$p < .025$

Note: Values are for event time only.

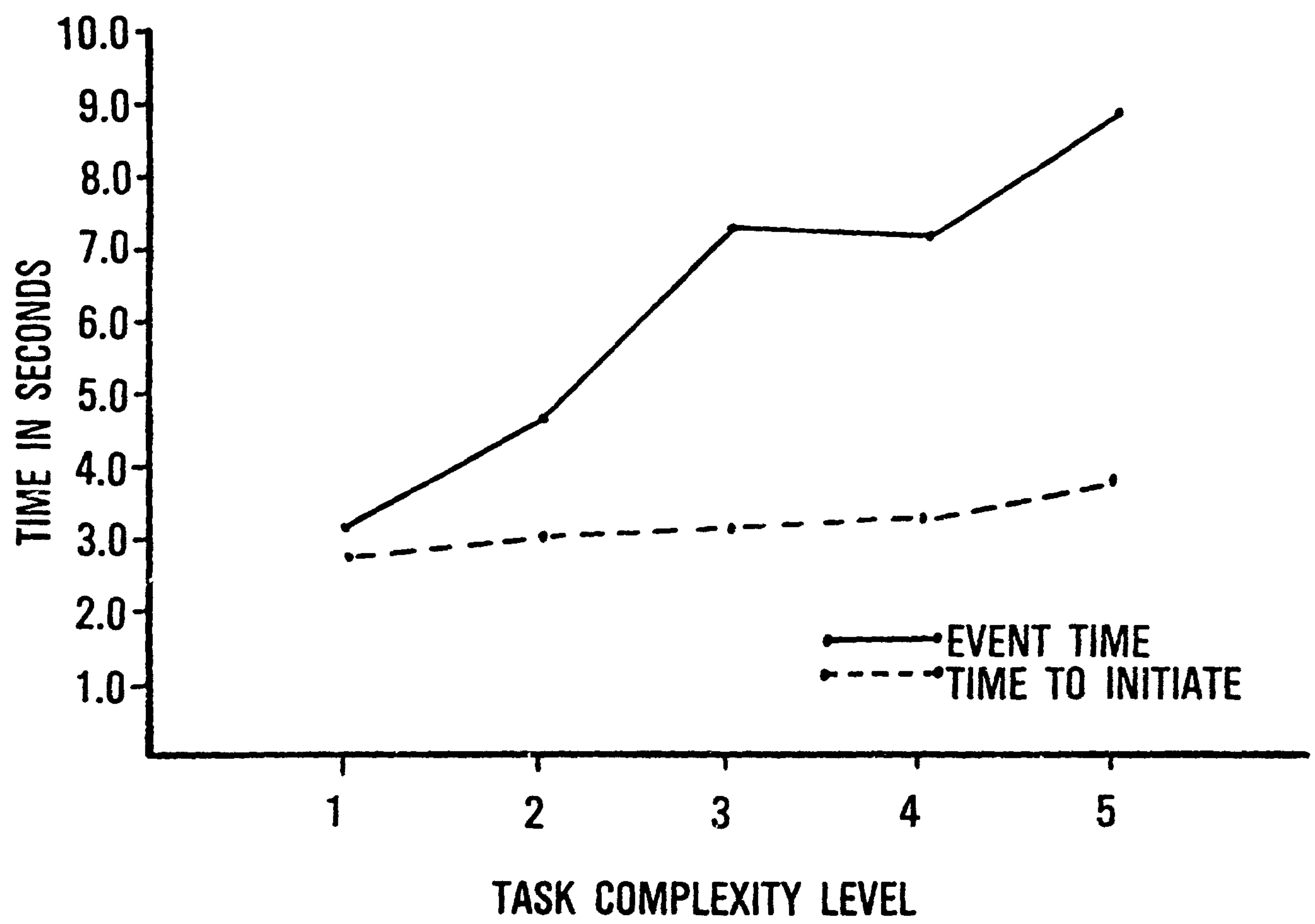

Figure 9. Mean time to initiate a task and event time by task complexity level.

Table 12

Results of the FIT Analysis
on the Interaction between Response Mode
and Task Complexity Level for Time to Initiate a Task

Comparison	d.f.	F	P
Single Task Manual vs Single Task Speech	1,827	118.874	$p < .025$
Dual Task Manual vs Dual Task Speech	1,827	27.519	$p < .025$

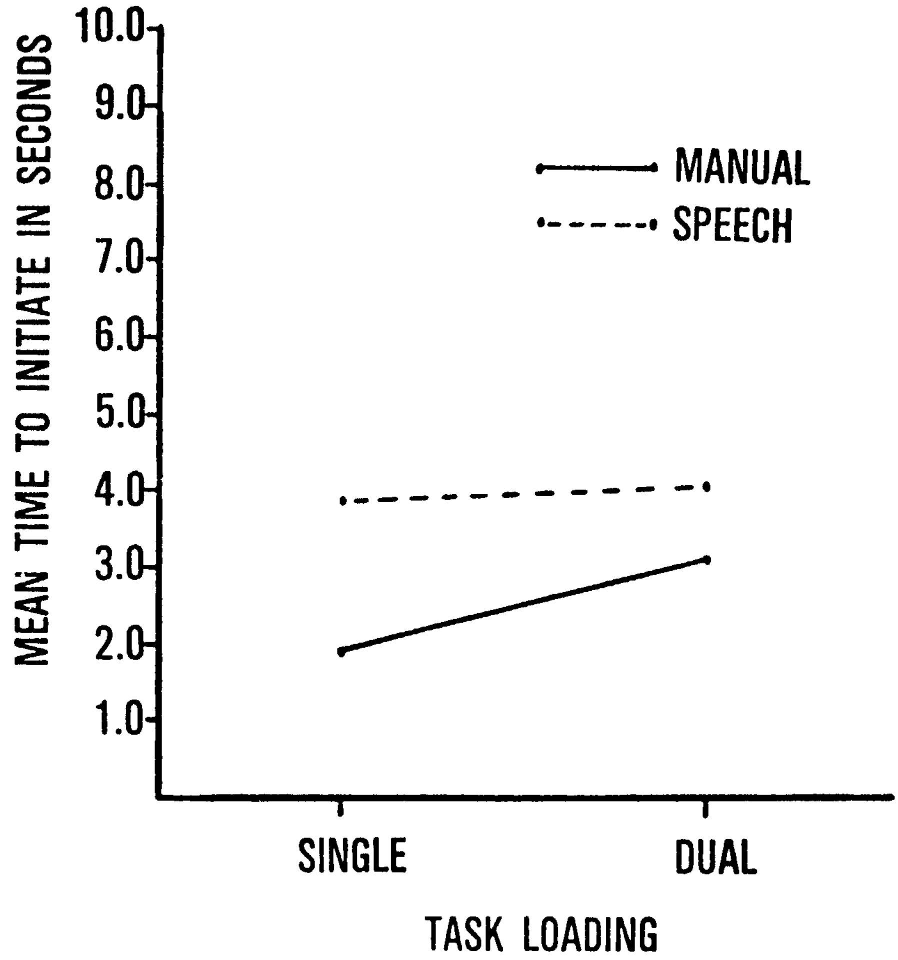

Figure 10. Mean time to initiate a task by task loading and response mode.

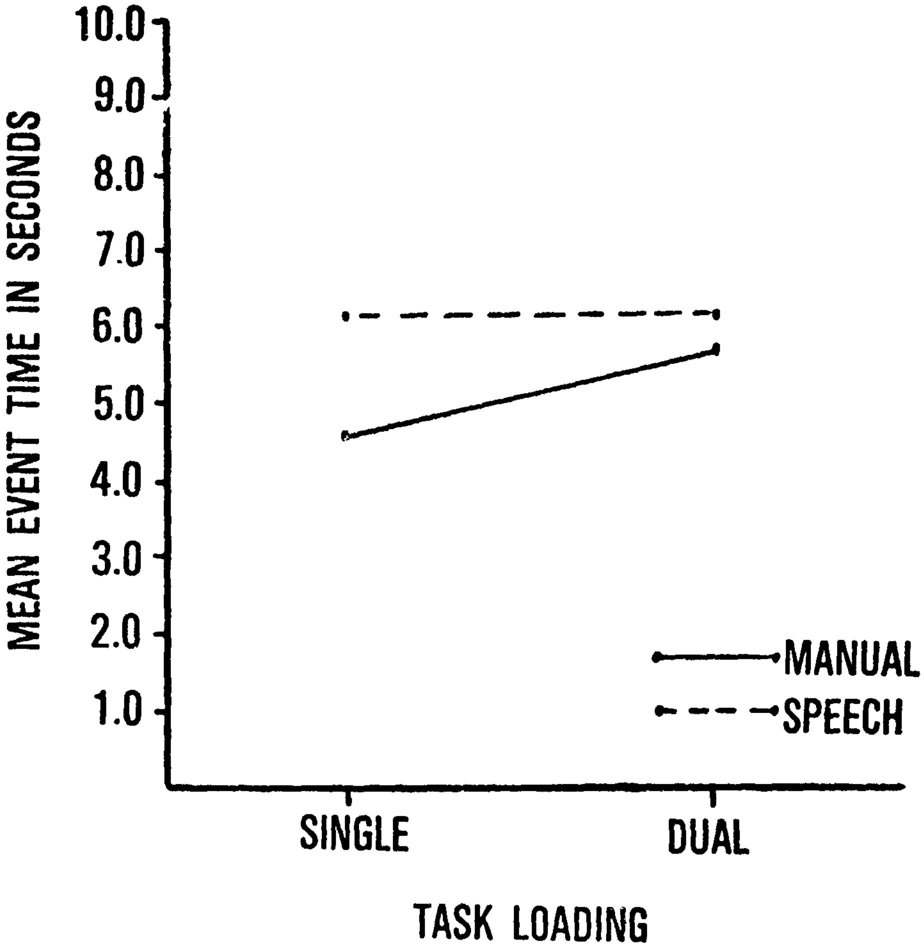

Figure 11. Mean event time by task loading and response mode.

than for the vocal response mode. The reason level 5 was not significant is probably due to the smaller sample size in the group. These results are shown in Table 13 and are depicted in Figure 12. For mean event time, the only significant differences found were for complexity levels 4 and 5. For these two complexity levels the manual response mode was significantly faster than the vocal response mode. The results of this analysis are shown in Table 14 and depicted in Figure 13. What is interesting about these results is that for complexity levels 1, 2, and 3, there were no significant differences between mean event time for the two response modes.

<u>Flying performance data analysis</u>

There were three dependent measures recorded to assess the pilots' flying performance: 1) vertical tracking error in pixels, (an error of one pixel was equal to .04 degrees error in the aircraft's desired position), 2) horizontal tracking error in pixels, and 3) airspeed deviation in knots. Vertical and horizontal tracking errors were recorded 10 times a second and computed as the difference between the aircraft's desired position (i.e., the flight director on the HUD) and the aircraft's actual position (i.e., the velocity vector). A root-mean-square (rms) error score was computed for each time segment involved in the analysis (i.e., an rms error score was computed for both pre-event and event time segments). Airspeed deviation was also recorded 10 times a second and was computed as the difference between the aircraft's desired airspeed (i.e., 420 knots) and the aircraft's actual airspeed. An rms error score was also computed for airspeed deviation for each time segment involved in the analysis.

A 2 x 2 x 5 repeated measures MANOVA (response mode x task loading x task complexity level) for unbalanced data was used to analyze the flight performance data and the results of this analysis are shown in Table 15. (The unbalanced procedure was again used since in using the raw

Table 13

Results of the FIT Analysis for Mean Time to Initiate a Task by Response Mode and Task Complexity Level

Task Complexity Level (Manual vs Dual)	d.f.	F	P
1	1,821	40.445	p < .025
2	1,821	20.850	p < .025
3	1,821	36.471	p < .025
4	1,821	21.740	p < .025
5	1,821	4.460	n.s.

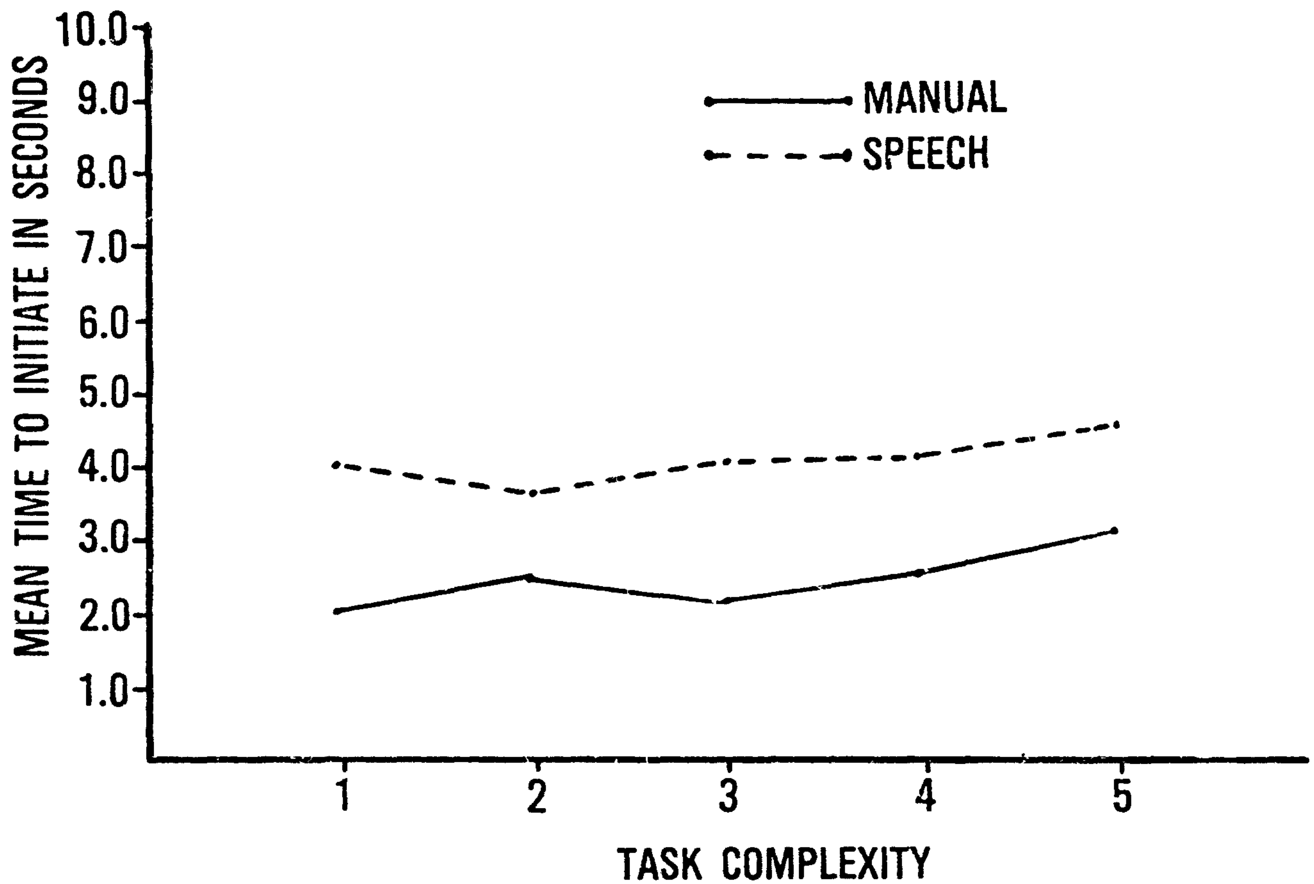

Figure 12. Mean time to initiate a task by task complexity level and response mode.

Table 14

Results of the FIT Analysis for
Mean Event Time by Response Mode and
Task Complexity Level

Task Complexity Level	d.f.	F	P
1	1,822	1.423	n.s.
2	1,822	.249	n.s.
3	1,822	7.757	n.s.
4	1,822	31.077	$p < .025$
5	1,822	27.014	$p < .025$

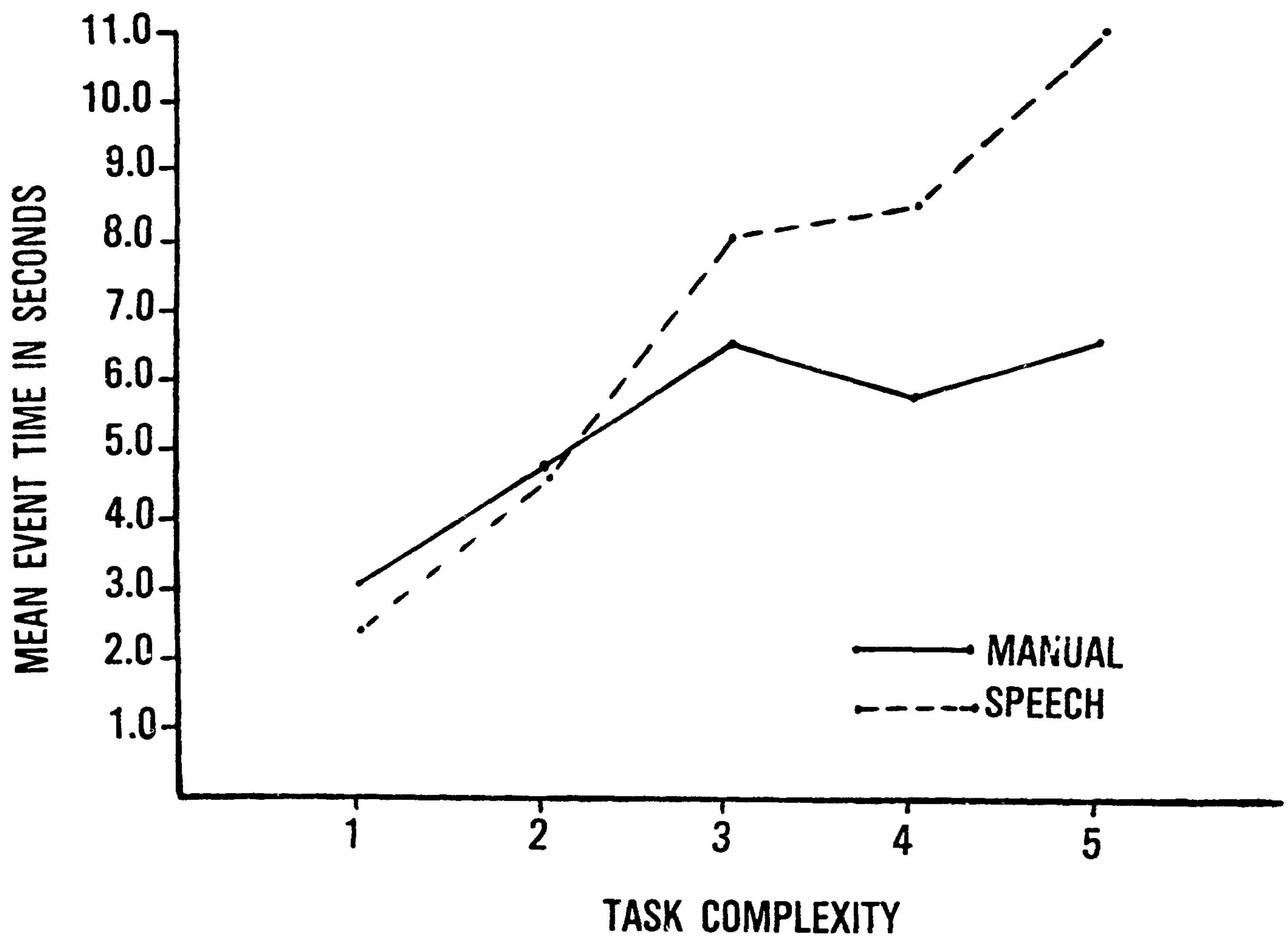

Figure 13. Mean event time by task complexity level and response mode.

Table 15

Results of the MANOVA on Flying Performance

Source	d.f.	F (Pillai's Trace)	P
Response Mode	3, 13	9.47	$p = .0014$
Task Loading	3, 13	17.01	$p < .0001$
Task Complexity Level	12, 180	1.42	$p = .1612$
Response Mode x Task Loading	3, 13	4.59	$p = .0211$
Response Mode x Task Complexity Level	12, 180	1.44	$p = .1533$
Task Loading x Task Complexity Level	12, 180	1.33	$p = .2055$
Response Mode x Task Loading x Task Complexity Level	12, 180	1.35	$p = .1937$

data task complexity level 5 had one third fewer observations than the other four task complexity levels.) Table 16 lists the means for the rms error scores. FIT was again used as a follow-up analysis to a significant MANOVA effect and these results are described in the following paragraphs.

<u>Response mode</u>. The FIT that examined the significant response mode main effect found that vertical tracking error was the only dependent variable which significantly contributed to the main effect, $F(1,830)=11.7$, $p < .017$. By comparing the means for vertical tracking error for the two response modes it can be seen that performance was better in the speech condition.

<u>Task loading</u>. For the task loading main effect the FIT revealed that both vertical tracking error and airspeed error significantly contributed to the main effect, $F(1,830)=25.0$, $p < .017$ and $F(1,828)=74.9$, $p < .017$, respectively. By comparing the means of these two variables it can be seen that performance on both measures was better in the single task condition.

<u>Response mode by task loading</u>. Figure 14 shows a plot of vertical tracking error, horizontal tracking error, and airspeed error by response mode and task loading. The results of the FIT that analyzed the significant interaction effect between response mode and task loading revealed some interesting results. First, there was no significant difference between single task performance for the two response modes, which would be expected; however, there was a significant difference between the two response modes for dual task performance and vertical tracking error was the only dependent variable that contributed significantly to this effect, $F(1,828)=15.4$, $p < .017$. (Vertical tracking error probably was the only significant dependent variable due to the fact the tracking task simulated terrain following flight, resulting in more vertical deviation in the flight director than horizontal deviation.) Second, a

Table 16

Mean rms Error Scores for Each Significant Effect

Effect	Mean rms Error VTE (Pixels)	HTE (Pixels)	ASE (Knots)
Response Mode			
Manual	12.54	6.75	2.94
Speech	10.62	6.15	3.23
Task Loading			
Single	10.18	6.10	2.42
Dual	12.97	6.81	3.75
Complexity Level			
1	10.38	6.54	2.80
2	11.71	6.30	3.00
3	12.54	6.80	3.34
4	11.46	6.27	3.21
5	12.25	6.18	3.03
Response Mode x Task Loading			
Single Task/Manual	10.58	6.26	2.43
Single Task/Speech	9.79	5.94	2.41
Dual Task/Manual	14.51	7.24	3.46
Dual Task/Speech	11.44	6.37	4.04

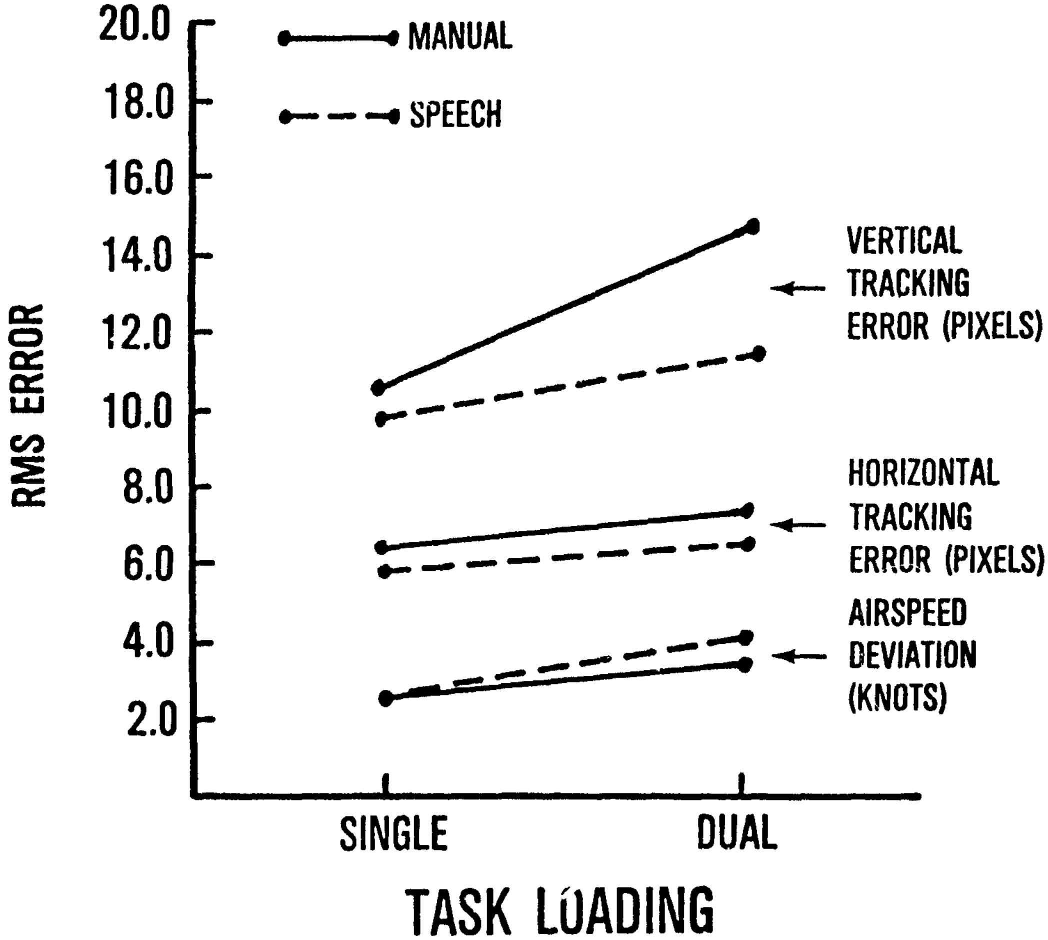

Figure 14. RMS error by task loading and response mode.

significant difference between single and dual task performance for the manual response mode was also found and both vertical tracking error and airspeed error contributed to the significant difference, $F(1,828)=25.2$, $p < .017$, and $F(1,826)=21.8$, $p < .017$, respectively. Finally, a significant difference between single and dual task performance for the vocal response mode was also found but airspeed error was the only dependent variable that significantly contributed to the effect, $F(1,826)=58.0$, $p < .017$. What is important about these results is that under dual task conditions tracking performance for the vocal response mode was better than for the manual response mode and was not significantly different than single task tracking performance. The airspeed error was greater for the vocal response mode but this finding was probably due to the fact that the switch which activated the speech recognizer was located on the throttle and had to be depressed when the pilot was talking to the speech recognizer. (It was assumed that this switch depression had minimal impact on the requirements for manual information processing resources.)

Error Data Analysis

For purposes of the error analysis, errors were defined as the number of switch hits or vocal commands made during a task minus the number required for the task. Thus, if 6 switch hits were required to complete a task and 6 switch hits were made the error would be 0. On the other hand, if 6 switch hits were required and 8 were made, there would be two errors counted. In this study 35 tasks accounted for a total of 78 errors, 47 errors for the manual response mode and 31 errors for the vocal response mode. It would appear that the vocal response mode had fewer errors, but these numbers are a little misleading. In fact, the vocal response mode probably would have had more errors due to misrecognitions by the speech recognizer but pilots often became confused on how to correct an error when a

misrecognition occurred and tasks had to be redone. For example, if a pilot said "OPTION 1" and the recognizer thought the pilot had said "OPTION 2" and activated the appropriate switch, the pilot usually would not know how to correct the mistake. In this situation, the task was terminated and repeated at the end of a flight.

The criteria used to determine when to repeat a task was when a pilot asked for instructions from the experimenter on how to correct an error made by the speech recognizer. This situation occurred a total of 42 times in the vocal response mode condition, whereas, only 5 tasks had to be repeated in the manual condition. The five manual tasks were also repeated as a result of pilot questions during the accomplishment of a task.

As a result of the problems just described, a general conclusion as to the error analysis is difficult to achieve. If the total number of errors were combined with the number of tasks that had to be repeated, the vocal response mode probably could be considered to have had more errors overall. Also, an appreciation is obtained for the fact that speech recognizer misrecognitions cause unexpected inputs that do not occur in a manual task unless a pilot presses the wrong switch.

Speech recognizer performance

In this study the speech recognizer had an overall recognition accuracy of 93.9 percent. When this number was corrected for pilot speaking errors, recognition performance increased to 95.0 percent overall. The number of misrecognitions for each word in the vocabulary is shown in Table 17.

Questionnaire Results

The results of the questionnaire are contained in Appendix D and were analyzed using a Kolmogorov-Smirnov goodness of fit test for a uniform distribution. There were several interesting findings from the questionnaire. First, pilots did not significantly favor the manually or vocally

Table 17

Number of Misrecognitions for Each Word in the Vocabulary

Word	Number of Misrecognitions	Number of Possible Entries	% Error
0	16	293	5.5
1	24	279	8.6
2	9	242	3.7
3	8	216	3.7
4	18	140	12.9
5	0	128	0.0
6	0	80	0.0
7	0	16	-
8	17	144	11.8
NINER	6	86	7.0
XRAY	0	32	0.0
YANKEE	4	36	11.1
CLEAR	1	7	-
UHF	10	74	13.5
TACAN	2	66	3.0
BAROMETER	0	64	0.0
FLY TO	14	210	6.7
DISENGAGE	0	64	0.0
COMM	0	64	0.0
IFF	7	199	3.5
NORMAL	1	65	1.5
MODE 1	0	64	0.0
MODE 3	18	87	20.7
OPTION 1	19	283	6.7
NOZE	0	32	0.0
TAIL	2	34	5.9
INTERVAL	11	75	14.7
QUANTITY	19	91	20.9
SINGLES	7	71	9.9
ENTER	21	617	3.4

activated MFC (Question #1) but did significantly prefer an MFC that could be operated manually or by voice (Question #2). Second, most pilots did describe the speech activated MFC as moderately easy to operate (Question #7) and said they would use a speech activated MFC quite often if future aircraft would include a speech recognizer (Question #9). Third, the pilots felt that the visual feedback provided on the HUD helped quite a bit (Question #10) and that speech recognition was a great advantage during head up flying (Question #11). Fourth, all the pilots agreed that speech recognition is a viable alternative for subsystem control in future aircraft (Question #8). Finally, probably the most enlightening aspect of the questionnaire was the comments made by the pilots. The comments are too lengthy to describe here, but they should be considered by any systems designer considering the use of speech recognition in the design of a system. One comment made by several pilots that should be given attention is that the response lag of the speech recognizer was too long and was frustrating during vocal input of data.

SECTION V

Discussion

Based on the results of this study the three hypotheses developed earlier can now be addressed. These hypotheses were 1) the vocal response mode would be more effective than the manual response mode, 2) the effectiveness of the response mode is related to the complexity of the task, and 3) the performance for the vocal response mode would be similar in both single and dual task conditions, whereas, performance for the manual response mode would deteriorate in dual task conditions. To bring the data just presented together to test these three hypotheses is a rather complex undertaking, but if a proper perspective is established the task becomes a little easier. For purposes of this discussion two different perspectives will be used: 1) the data entry is the primary task and 2) flying the aircraft is the primary task. Depending on which perspective you take, the results of this study can look quite different.

If it is assumed that data entry was the primary task, it can probably be concluded that manual data entry was most effective. This conclusion is based on the fact that both mean time to initiate a task and mean event time were shorter for the manual response mode. However, these times do include the response lag for the speech input mentioned earlier which would tend to put the vocal response mode at a disadvantage. If 400 msec. were subtracted from the event time for each vocal command entry (e.g., if there were 3 entries, 1.2 seconds would be subtracted), the mean event time for the vocal response mode condition would go from 6.29 seconds to 4.60 seconds which is faster than the 5.21 seconds for the manual response mode.

Even with the response lag in the vocal condition, the mean event times were not significantly different for the two response modes for task complexity levels 1, 2, and 3. This finding supports the research summarized earlier that

showed speech performs better when the data to be entered is words or phrases, whereas the manual response mode performs better when the data requires character by character input. For complexity levels 1, 2, and 3, the data to be entered tended more toward words, but for complexity levels 4 and 5 the data tended more toward digits. The same point about the response lag for the vocal response can be made again here. If 400 msec. were subtracted from the event time for each vocal command the mean event times for each complexity level would decrease. But for complexity level 5 for example, the manual response mode would still be faster with a mean of 6.74 seconds as compared to the corrected mean of 8.39 seconds for the vocal response mode. Therefore, the manual response mode would still be more effective than the vocal response mode for the more complex tasks; however, the vocal response mode would still be just as effective, if not more so than the manual response mode for the less complex tasks.

As far as the third hypothesis is concerned, that the vocal response mode would reduce competition for information processing resources in the dual task condition, it appears that it is true. Performance for the vocal response mode remained similar in both single and dual task conditions, whereas, the performance for the manual response mode declined in the dual task condition when there was competition for manual resources from the flying task. Based on these results, a parallel processing model of human information processing gains support since the capability to enter data by voice reduced the competition for information processing resources.

If the perspective is taken where flying the aircraft is the primary task, the data takes on a little different slant. First of all, flying performance becomes the criteria, whereas event time was the criteria earlier. If the effect of the two response modes on flying performance is analyzed, it is clear than the vocal response mode had

the least impact on performance and is more effective for data entry. This conclusion is based on the fact that tracking performance for the vocal response mode was similar in both single and dual task conditions, whereas, tracking performance deteriorated for the manual response mode under dual task conditions.

These results would again support a parallel processing model of human information processing. Under dual task conditions, when there was competition for manual resources, flying performance deteriorated in the manual condition but performance for the vocal condition remained relatively constant since there was little competition for resources. These results would support the idea that humans can process information in parallel to accomplish concurrent tasks.

In putting both perspectives together, it appears that the pilots adopted a tradeoff strategy dependent upon their current workload. In the situation when data entry was done manually the pilots would concentrate most of their information processing on the data entry task and flying performance would suffer; however, when data entry was done vocally the pilots could still concentrate on the flying task and enter data while maintaining good flight control.

SECTION VI
Conclusions

In summarizing this discussion as to whether the manual or vocal response mode is more effective in accomplishing aircraft systems tasks, it can be said that a tradeoff occurs that is dependent upon the current workload of the pilot. In an operational aircraft the pilot's workload would vary as a function of the mission phase. Therefore, in a low level terrain following phase when flight control of the aircraft is critical, the vocal response mode is probably the most effective alternative since it has the least impact on flight performance. In this situation the pilot would concentrate his information processing resources on the flying task and enter data by voice. In a cruise phase, when flight control is not that critical, the manual response mode is probably more effective since the tasks can generally be accomplished quicker with fewer errors. In this situation the pilot would concentrate his information processing resources on data entry. Therefore, the issue is not whether to implement speech or manual input, but since the effectiveness of either response mode is a function of the pilot's current workload, it appears that both response alternatives should be provided to the pilot for his selection. This conclusion is also supported by the questionnaire results in which the pilots overwhelmingly favored the implementation of both vocal and manual control. In future studies perhaps these results could be verified or improved upon with a speech recognizer that is not handicapped by an unnecessary response lag as contained in this study.

SECTION VII

Future Research

From the process of conducting the current study, several subjective conclusions can also be reached as to some important topics that future research in this area could address.

First, from the comments made by the pilots, it became apparent that the response lag of the speech recognizer used in this study was unacceptable. Future research should address the question as to what is an acceptable response time for specific speech recognition applications. In some applications the response time may not be critical, but in military airborne applications where workload tends to be high, response time can be very critical.

Second, the speech recognizer used in this study had an overall recognition accuracy of 95.0 percent; however, the recognition accuracy dropped as low as 87.5, 88.6, and 89.1 for some pilots. These values are unacceptable. Speech recognition algorithms still need to be improved upon to attain high recognition accuracies for all pilots, not just high averages.

Third, research should be conducted to determine just what error rate is acceptable for airborne and other applications of speech recognizers.

Fourth, some words in this study had higher error rates than other words in the vocabulary. Future research could address size and confusibility issues of a vocabulary that would be required in an airborne environment . The present study only used a small subset of the potential vocabulary.

Finally, since speech recognition would be difficult in an airborne environment, due to noise, vibration, g's, etc., future research could address the possibility of tailoring recognition algorithms and vocabularies as a function of the mission phase. For example, in an air-to-air combat phase when the pilot is under a lot of stress and g's, perhaps the

speech recognizer could be programmed to switch to a recognition algorithm and vocabulary list specifically designed for that phase. Hopefully, recognition tailoring could help improve recognition accuracies in adverse situations when a high payoff from speech recognition could be realized.

This is just a partial list of the several possible areas for future research in speech recognition. Even though problems still exist for speech recognizers, research such as the present study points to advantages that support the push for applications of speech recognition in systems. The most obvious and important advantage being a more efficient distribution of workload across available output modalities. However, systems designers must be careful as to how speech recognition is applied. Speech recognition should not be used because it is a novel technology. Rather, a prudent designer would only use speech recognition when it would truly be more effective than alternative approaches.

Appendix A
Head-Up Display Format

The purpose of this section is to describe the HUD format used in this study in more detail. The HUD symbology was generated by a RAMTEX symbol generator and presented on a raster scan CRT with a resolution of 320 x 240 pixels. The pilot viewed the HUD symbology through an optical combiner glass on top of the glare shield (Figure 3). The following paragraphs describe the actual symbology.

The aircraft velocity vector was represented by a flight path marker (FPM) which denoted the point toward which the aircraft was flying at all times. The FPM moved horizontally and vertically based on the pilot's control inputs, but was not roll-stabilized to show bank angle. Rather, the flight path scales and their associated numbers were roll-stabilized and rotated to the appropriate bank angle.

The flight director symbol indicated horizontal and vertical steering error with respect to the flight path marker. The X, Y commands to position the flight director symbol were such that the pilot flew the flight path marker to the flight director by steering the aircraft in pitch and/or bank angle, i.e., the flight director was moved by the software to the flight path marker when it received the proper control signals. The forcing function used to introduce error in the flight director position was based on a gust model that simulated flight in windy conditions. The error to be entered was computed in the roll, pitch, and yaw axis and was based on a random number generator. The error values were computed at a rate of 20 Hz but were damped by averaging three values before they displaced the flight director. This damping function resulted in smoother displacement of the flight director that resembled actual flight in windy conditions. The wind conditions simulated would be considered mild turbulence.

The horizon and flight path angle lines of the flight path scale represented the horizon and each five degrees of flight path angle (FPA) between plus and minus 90 degrees. Positive FPA was presented as solid lines and appeared above the horizon line. Negative FPA was presented as dashed lines and appeared below the horizon line. The five degree increments were numbered on either end of the FPA lines. A minus sign preceded the numbers for negative angles.

The airspeed, heading, and altitude scales were not roll-stabilized. The airspeed and altitude scales were vertical and appeared on the left and right sides of the display, respectively. The heading scale was horizontal and appeared at the top of the display. The airspeed scale was graduated in 25 knot increments and numbered each 50 knots. An exact readout of current airspeed was presented in the window in the center of the scale. The readout changed whenever the airspeed changed by one knot.

Barometric altitude was displayed on the altitude scale on the right side of the HUD. The scale was graduated in 250 foot increments numbered each 500 feet. The total range of the altitude scale was from minus 1,000 feet to plus 99,999 feet with 1,500 feet in view at all times. An exact readout of the altitude was provided in the window in the center of the scale. The readout changed whenever the altitude changed by 1 foot.

The heading scale was displayed at the top of the HUD. Forty degrees were in view at all times, graduated in five-degree increments, labeled with two-digit numbers every ten degrees. Total heading scale range was 360 degrees. The aircraft magnetic heading was displayed to the nearest degree in the window.

Vertical velocity was displayed in the upper right corner of the HUD (above altitude scale) in digital form with the readout changing in 1 foot per minute increments over a range of 0 to 9,999 feet per minute. A caret (^) indicated vertical velocity direction, i.e., up or down.

Appendix B
Experimenter's Console

The experimenter's console was equipped with CRT displays and status light matrices that provided the experimenter with the capability of monitoring the displays in the simulator and the pilot's actions. A layout of the experimenter's console is shown in Figure B1. The following list specifies the functions allocated to each piece of equipment on the console used in the present study. Each letter refers to the notations used on the layout.

A = Status display; presented flight and task event information

B = Not used

C = Not used

D = Repeater display of HUD

E = Not used

F = Not used

G = Status panel lights; each status light stayed lit as long at the corresponding switch in the cockpit was activated.

H = Master power switch for facility

I = Abort switch for McFadden flight control system

J = Interphone options (Note: the pilot's mike was always hot.)

K = On/off switch for interphone system

L = Switch enabled communication between two experimenters

M = Switch enabled experimenter/pilot communication

N = Switch enabled experimenter/computer personnel communication

O = Switch enabled communication between experimenters, pilot, and computer personnel

P = Volume control for headset

Q = Voice recorder options

R = Run switch for voice recorder

S = Pause switch for voice recorder

T = Reset switch for McFadden flight control system

U = Pre-event switch; activation initiated fifteen seconds of flight data recording

V = Event marker switch; activation started recording of task event data and unlocked MFC

W = Mission complete switch (guarded); activation initiated the computerized data reduction procedures

X = Run switch for simulation

Z = Keyboard unlock switch; activation unlocked MFC in those task events where recording terminated after the pilot entered incorrect legal digits

A1 = Indicated whether tape recording was continual or voice activated

A2 = Hold switch for simulation

A4 = Not used

A5 = Task abort switch (guarded); activation terminated recording of task event data and initialized system for next task event.

A6 = Repeater display of MFC

A7 = Repeater display of upper left CRT

A8 = Speech Recognizer status; indicated whether or not the speech recognizer recognized a word properly.

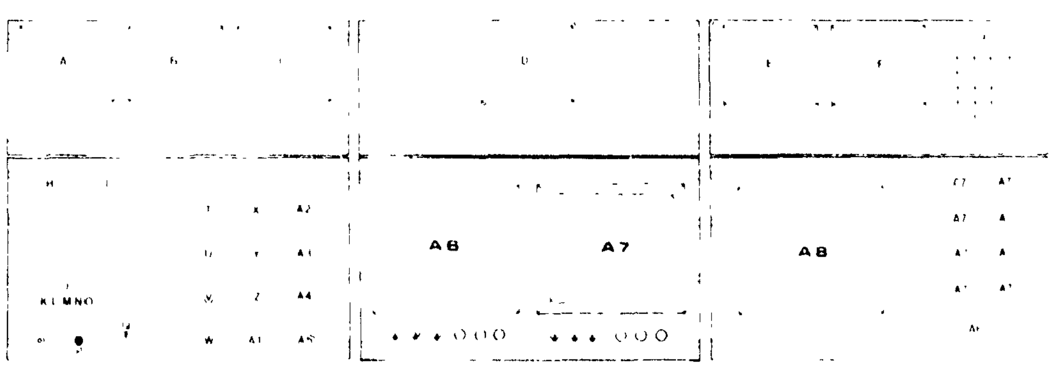

Figure B1. Experimenter's Console.

Appendix C

Experimental Matrix

MATRIX	PILOT	TREATMENT	MISSION	MATRIX	PILOT	TREATMENT	MISSION
1	1	2	2	33	9	2	1
2	1	1	1	34	9	4	2
3	1	3	3	35	9	1	3
4	1	4	4	36	9	3	4
5	2	1	2	37	10	4	3
6	2	4	1	38	10	3	1
7	2	2	3	39	10	2	2
8	2	3	4	40	10	1	4
9	3	3	1	41	11	1	1
10	3	2	4	42	11	2	3
11	3	4	2	43	11	3	2
12	3	1	3	44	11	4	4
13	4	4	3	45	12	3	3
14	4	3	2	46	12	1	1
15	4	1	4	47	12	4	2
16	4	2	1	48	12	2	4
17	5	1	1	49	13	4	4
18	5	3	3	50	13	1	3
19	5	2	2	51	13	2	1
20	5	4	4	52	13	3	2
21	6	3	3	53	14	1	4
22	6	4	1	54	14	3	3
23	6	1	2	55	14	4	1
24	6	2	3	56	14	2	2
25	7	2	4	57	15	2	3
26	7	1	3	58	15	4	2
27	7	4	2	59	15	3	4
28	7	3	1	60	15	1	1
29	8	4	3	61	16	3	1
30	8	2	1	62	16	2	4
31	8	3	2	63	16	1	2
32	8	1	4	64	16	4	3

Note: Treatments

1 = Manual

2 = Speech

3 = Manual and Flying

4 = Speech and Flying

Appendix D

Mission Script

Data Mission #2

TASK 1

Experimenter: Boxcar 11, change your TACAN to 115 X-ray.

Boxcar 11: Roger

TASK 2

Experimenter: Boxcar 11, squawk normal.

Boxcar 11: Roger

Experimenter: I read your squawk.

TASK 3

Experimenter: Boxcar 11, change your barometer setting to 30.02.

Boxcar 11: Roger

TASK 4

Experimenter: Boxcar 11, change your UHF frequency to 348.0.

Boxcar 11: Roger

TASK 5

Experimenter: Boxcar 11, choose option 1 and change your drop mode to singles.

Boxcar 11: Roger

TASK 6

Experimenter: Boxcar 11, choose weapon option 1 and select tail fuzing.

Boxcar 11: Roger

TASK 7

Experimenter: Boxcar 11, choose option 1 and change your quantity to 18.

Boxcar 11: Roger

TASK 8

Experimenter: Boxcar 11, change your mode 3 squawk to 6500.

Boxcar 11: Roger

Experimenter: I read your squawk.

TASK 9

Experimenter: Boxcar 11, engage your fly to and enter waypoint number 3.

Boxcar 11: Roger

TASK 10

Experimenter: Boxcar 11, choose weapon option 1 and change your interval to 50 ft.

Boxcar 11: Roger

TASK 11

Experimenter: Boxcar 11, select fly to and enter waypoint number 4.

Boxcar 11: Roger

TASK 12

Experimenter: Boxcar 11, change your mode 1 squawk to 62.

Boxcar 11: Roger

TASK 13

Experimenter: Boxcar 11, disengage fly to.

Boxcar 11: Roger

Appendix E

An Evaluation of Speech Recognition for the Control of Aircraft Subsystems

Final Debriefing Questionnaire

We are doing a study to determine the effectiveness of using computer speech recognition to control non-critical aircraft subsystems. The results of this study will help determine what ground rules should be used in designing future generation crew stations. One of the best sources of information in terms of improving avionics designs and avoiding mistakes is to talk directly with pilots who fly operational aircraft. Your candid opinion will help a great deal in achieving this goal.

Most of the questions can be answered with a single check (✓) (only check one block); however, any additional comments that you provide will be very helpful. Please be as specific as you can.

PERSONAL DATA

NAME ____________________

ORGANIZATION/OFFICE SYMBOL ____________________

PHONE ____________________

JOB TITLE ____________________

TOTAL FLYING TIME $\overline{X}$ = 2241 hours

TOTAL JET TIME $\overline{X}$ = 1584 hours

AIRCRAFT FLOWN	TIME IN TYPE

TOTAL YEARS RATED $\overline{X}$ = 9.25 years

AGE $\overline{X}$ = 33.4 years

WHAT AIRCRAFT ARE YOU CURRENTLY QUALIFIED IN? ____________________

1. Considering all the tasks completed in this simulation, compare manual operation of the multifunction control (MFC) with the speech activation of the MFC. Check (✓) the appropriate box.

Manual Much Better Than Speech	Manual Slightly Better Than Speech	Equal	Speech Slightly Better Than Manual	Speech Much Better Than Manual
2	4	2	4	4

$\underline{D}$ = .1 (n.s.)

Subject 1: I felt more confident in manual operation.

Subject 2: Speech is better since you can monitor it on the HUD while flying.

Subject 3: Control improves greatly when you do not have to look down. (In the speech condition.)

Subject 4: Manual operation as set up in your simulation is a quantum leap above operation aircraft. Voice is a quantum leap above that.

Subject 5: Speech would be much better if you weren't worried about proper pronunciation each time.

Subject 6: Manual is slightly better only because of the difficulty of making sure your speech is identical to the coded speech and the terminology used was not similar to that used in the F-4.

Subject 7: Some tasks didn't lend themselves well to voice, i.e., altimeter changes. Some were pleasant -- Option 1 list changes.

Subject 8: The manual required pilot to drag eyes from HUD but voice required repetition and fixation on MFC display or HUD to confirm proper action.

Subject 9: The actual task completion would be about the same, but the location of the control heads for manual inputs were bad. As a pilot becomes more familiar with the control heads, the manual and voice activation would be about the same.

Subject 10: Manual was more accurate and inputs were taken with less time and trouble. Speech however, was much more convenient and made the flying task easier. When the programming reduces errors in recognition I would check this box far right (speech much better than manual).

Subject 11: The speech could have been much better if the feedback could be faster.

Subject 12: Manual more reliable at this stage.

Subject 13: Speech activation required very slow verbal input. More time was required than with a manual input. With practice though, this could be reduced.

Subject 14: If there were fewer constraints on the vocabulary/syntaxing, it would rate "much better."

Subject 15: The actual operation is quicker and I had fewer errors with the manual system, but for some particular tasks, the manual operation caused greater distraction and less accurate flight control.

Subject 16: Speech mode allowed for more attention to aircraft control.

2. Rank order the following in terms of which method of controlling non-critical aircraft subsystems you would like to see in future cockpits. (1 = first choice, 2 = second choice, etc.)

_____ Conventional control heads (A-7, F-15, F-4, etc.)

_____ Manually activated MFC

_____ Speech activated MFC

_____ An MFC that can be activated manually or by speech

COMMENTS:

	1st	2nd	3rd	4th	Choice
CONV	0	1	4	10	$\underline{D}$ = .433 ($p < .05$)
MAN	2	5	7	1	$\underline{D}$ = .180 (n.s.)
SPEECH	0	6	5	4	$\underline{D}$ = .250 (n.s.)
MAN/SP	14	2	0	0	$\underline{D}$ = .680 ($p < .05$)

Subject 1: No comment

Subject 2: No comment

Subject 3: I like the flexibility of both ways. May be due to lack of experiential confidence in speech control.

Subject 4: Maybe it's my psychological resistance to advanced technology, but I'm not sure I'd feel totally comfortable with speech activated controls with no back up manual system.

Subject 5: No comment

Subject 6: No comment

Subject 7: Each airplane has systems that lend themselves to voice -- perhaps each is different based on A/C capabilities. F-16 multifunction displays are a good target effort.

Subject 8: No comment

Subject 9: I rated the speech activated only MFC last since in an environment of radio transmissions the pilot really cannot just sit there and talk to his airplane. Such an environment would be while being vectored by a FAC, AWACs or GCI site. I also wonder about the voice recognition capabilities of the system when the pilot has been breathing 100% oxygen for a while and talking and his voice, as a result, gets pretty garbaged up.

Subject 10: No comment

Subject 11: Unless the speech recognition can be 100% error free there are definitely things which must still be done manually.

Subject 12: No comment

Subject 13: In critical parts of flight this method could be very beneficial in safety and performance capabilities -- increasing both.

Subject 14: No comment

Subject 15: No comment

Subject 16: No comment

3. Is there any function or functions that you feel are so critical that they should not be activated by voice?

Subject 1: Weapon Release.

Subject 2: Bailout--Throttle and stick control.

Subject 3: Not if you have a separate voice control keying switch.

Subject 4: Basic flight controls. Emergency systems (engine shutdown, bailout, fuel shut-off, etc.)

Subject 5: Ejection system, aircraft control, nuclear systems.

Subject 6: Air-to-air weapons selection should not be activated by voice because of the "emotional" changes that occur in a pilots voice intensity, etc. during a combat engagement.

Subject 7: Ejection

Subject 8: Depending upon the recognition rate--weapons release (i.e. needs 100%). Gear extension/retraction, flop retraction/extension, speed--break actuation.

Subject 9: Emergency procedure items.

Subject 10: I would never try a power, altitude, heading, or airspeed setting. These functions can be set in some flight directing auto pilots and I think they should not be voice set or activated.

Subject 11: Weapons release. Anything which would cause irreversible effects e.g., gear down at 450K.

Subject 12: Weapons stations.

Subject 13: Ejection. AFCS auto pilot.

Subject 14: Nuclear weapon arming and release.

Subject 15: Critical emergency systems such as fire handles, ejection handle, etc.

Subject 16: No comment.

4. Rate the acceptability of a manually activated MFC to control aircraft subsystems.

UNACCEPTABLE	BAD	SATISFACTORY	GOOD	OPTIMUM
0	0	3	12	1

$\underline{D} = .413$ ($p < .05$)

COMMENTS:

Subject 1: No comment

Subject 2: No comment

Subject 3: No comment

Subject 4: No comment

Subject 5: No comment

Subject 6: Logic patterns such as changing squawk to normal would appear more logical to me with a squawk button instead of COMM, then, IFF, then to normal.

Subject 7: We can do better! Upfront controls are a step in this direction. The difficulty is getting lost in pages and lists.

Subject 8: No comment

Subject 9: It was good except for the location of the control heads.

Subject 10: The design and multifunction capability reduces the clutter of switches and boxes. It seemed to be an efficient effective way to control.

Subject 11: No comment

Subject 12: No comment

Subject 13: Much better than present A-7 arrangement. Quicker and more organized. Safer. Voice would be the only thing better other than mind control (mental telepathy).

Subject 14: Rating would be highly dependent on logic structure used and mechanization of the display (touch panel, etc.).

Subject 15: No comment

Subject 16: The manually activated MFC is very good. By centralizing all subsystem controls, it lowers the pilot's workload. It does however, keep the workload high enough to keep the pilot active and thinking.

5. Rate the acceptability of a speech activated MFC to control aircraft subsystems.

UNACCEPTABLE	BAD	SATISFACTORY	GOOD	OPTIMUM
0	0	5	8	3

$\underline{D}$ = .4 ($p < .05$)

COMMENTS:

Subject 1: No comment

Subject 2: No comment

Subject 3: No comment

Subject 4: Optimum, however, I feel that a manual system for backup and possible higher versatility is required.

Subject 5: No comment

Subject 6: Very hard to continually speak during a simulator mission just as you did when coding the computer.

Subject 7: We need to explore which function can use voice. Some things are best left manual (altimeter changes). I think efforts in weapons loading and selection are excellent candidates.

Subject 8: No comment

Subject 9: Since it lagged behind what I said, I sometimes wondered if it heard me. Would such a system require training before each flight or would they build data bases of all the pilot's voices with the vocabulary required to operate subsystems?

Subject 10: The current system allows quite a few errors and at times after the third time trying to get a word recognition I wanted a manual MFC. However, when the voice recognizer worked it greatly reduced the workload and made it easier to fly.

Subject 11: No comment

Subject 12: No comment

Subject 13: Would be optimum if 1) faster input capabilities and 2) voice recognition could be programmed quickly.

Subject 14: If recognition accuracy is sufficiently high and speech rate is unconstrained. Otherwise rating is "good."

Subject 15: No comment

Subject 16: I feel that the speech mode, although superior to the manual MFC for controlling subsystems, is probably a less desirable system. By making controlling subsystems so easy, it could encourage laziness in the pilot by reducing his workload too much.

6. For each task (i.e., UHF, IFF, etc.) compare the manually activated MFC with the speech activated MFC.

	Manual Much Better Than Speech	Manual Slightly Better Than Speech	Equal	Speech Slightly Better Than Manual	Speech Much Better Than Manual	D
UHF CHANGE	1	4	4	4	3	.14
TACAN CHANGE	0	4	4	5	3	.20
BAROMETER CHANGE	1	4	3	5	3	.14
IFF NORMAL	1	0	0	9	5	.53
IFF MODE 1 CHANGE	1	4	4	4	3	.14
IFF MODE 3 CHANGE	1	3	6	2	3	.13
FLY TO ENGAGE	0	3	6	3	4	.21

FLY TO DISENGAGE	0	3	6	3	4	.21
FLY TO CHANGE	0	4	5	3	3	.20
WEAPON DROP MODE CHANGE	0	3	3	6	4	.23
WEAPON FUZING CHANGE	0	2	3	6	5	.29
WEAPON INTERVAL CHANGE	0	3	4	5	4	.21
WEAPON QUANTITY CHANGE	0	3	4	5	4	.21

COMMENTS:

Subject 1: IFF Mode 1 and 3 Changes: I had a lot of trouble with these two. The speech pattern that I was using caused a problem. I changed the spacing between Mode and 1/3 which caused an error in recognition.

Subject 2: No comment

Subject 3: Speech functioned better in the flying environment. It was also easier in the nonflying mode.

Subject 4: UHF, TACAN, BAROMETER CHANGE, IFF CHANGE, all involve changing a set of digits. This is where speech activation shines.

Subject 5: No comment

Subject 6: No comment

Subject 7: Fly to Modes--With upfront controls these functions are only "one button" away. I don't think these are prime candidates. Weapons--Again, I feel these are excellent choices. COMM/NAV--Could be good choices; I thought there was a lack of rapid positive feedback on digit selection. I'm sure this could be improved. IFF M/C--I confused this with the "C" Mode of IFF. (Altitude Reporting).

Subject 8: No comment

Subject 9: I felt they were for the most part about equal. For the most part, I liked the speech systems better (slightly) than the manual system because I could just use the HUD. I felt the control head location for the manual system was poorly placed. However, as time progressed I thought the tasks could be accomplished with the same ease, whether speech or manually activated.

Subject 10: The items where speech was much better than normal were due to a good recognition capability. When I

had to change an item with a number I wasn't sure it would work the first time.

Subject 11: All of these items are reversible without damage or harmful effects.

Subject 12: No comments

Subject 13: Dependent upon the mental sorting (position of push buttons) required to locate MFC function operations.

Subject 14: Would suggest using the word "ALTIMETER" in place of "BAROMETER" to better fit natural syntax.

Subject 15: No comment

Subject 16: No comment

7. How easy was it for you to use the speech activated MFC?

VERY DIFFICULT	MODERATELY DIFFICULT	NO OPINION	MODERATELY EASY	VERY EASY
0	3	0	10	3

$\underline{D}$=.41 ($p < .05$)

COMMENTS:

Subject 1: I think once I got used to the voice system through constant use it would be easier to use and be a big help in aircrew workload.

Subject 2: No comment

Subject 3: No comment

Subject 4: Very easy would have been my ranking after a few more training missions.

Subject 5: With additional practice it could be quite easy.

Subject 6: No comment

Subject 7: No comment

Subject 8: No comment

Subject 9: No comment

Subject 10: This is due to the slow speech pattern that has to be used. Also, some numbers like 2 and 8 seemed to be particularly difficult for the processor to recognize. I did not have any real difficulty in getting used to the acquisition system.

Subject 11: Except after a while the word "enter" could not be recognized.

Subject 12: No comment

Subject 13: More time with the system would improve this.

Subject 14: Using the same logic for both manual and speech mode transitioning easier, but resulted in less than optimal logic for the speech.

Subject 15: No comment

Subject 16: No comment

8. Do you feel speech recognition is a viable alternative for the control of aircraft subsystems in future generation aircraft?

NO ___0___ YES ___16___

COMMENTS:

Subject 1: No comment

Subject 2: No comment

Subject 3: Provided you can work out the recognition for voice changes due to medical and environmental reasons (Gs and stress) and you can filter out the changing background noise.

Subject 4: Especially for high frequency subsystem changes involving a change in digits.

Subject 5: No comment

Subject 6: No comment

Subject 7: No comment

Subject 8: No comment

Subject 9: No comment

Subject 10: Allows much better aircraft control or at least an easier task load.

Subject 11: No comment

Subject 12: No comment

Subject 13: With greater workload in the cockpit, this would be a task reducer/simplifier.

Subject 14: No comment

Subject 15: No comment

Subject 16: No comment

9. If a future generation aircraft contained both speech and manually activated subsystems and either speech or manual activation could be used for any subsystem, how often would you use the speech activated control mode?

NEVER	VERY LITTLE	SOME	QUITE OFTEN	ALWAYS
0	0	3	12	0

$\underline{D}=.4$ ($p < .05$)

COMMENTS:

Subject 1: No comment

Subject 2: No comment

Subject 3: No comment

Subject 4: No comment

Subject 5: No comment

Subject 6: No comment

Subject 7: Premature question--If it makes today's tasks easier than always, if not, very little.

Subject 8: No comment

Subject 9: I feel this would be the best of both worlds. As for the amount of times I would use it, I can honestly answer only SOME because it would be a function if where I

was at in a given mission. I think I would only use the voice activated system in non-critical phases of flight.

Subject 10: With improvements made in the deficiencies I noted earlier, I would probably use speech activated 99% of the time. I would like the manual as a backup and cannot give any specifics of when I would use it. High tasks loads such as departure and approach should be significantly reduced with speech.

Subject 11: No comment

Subject 12: No comment

Subject 13: Dependent upon Initialization/Voice Recognition input time requirements.

Subject 14: No comment

Subject 15: No comment

Subject 16: Using the speech activated control system would depend on phase of flight and on how much easier speech activation would be compared to manual activation.

10. How helpful was the feedback provided on the HUD as to what word was recognized or that the speech recognizer did not understand you?

NOT HELPFUL AT ALL	HELPED VERY LITTLE	HELPED SOME	HELPED QUITE A BIT	ALWAYS HELPED
0	0	0	9	7

$\underline{D} = .6$ ($p < .05$)

COMMENTS:

Subject 1: No comment

Subject 2: No comment

Subject 3: No comment

Subject 4: No comment

Subject 5: No comment

Subject 6: No comment

Subject 7: A little slow for UHF/TACAN feedback changes. It's important that the pilot sees right away that the UHF/COMM change was made; so he can get on with other work of more importance.

Subject 8: No comment

Subject 9: It helped, but due to the slowness in reaction time, I wondered if it heard me. Maybe if while it was finding out what the word was if it would flash an asterisk or something at me to let me know it heard would be comforting.

Subject 10: No comment

Subject 11: If only it could be faster!

Subject 12: No comment

Subject 13: Heads up as much as possible.

Subject 14: No comment

Subject 15: Even very good for manual.

Subject 16: No comment

11. How advantageous do you feel speech recognition is during head up flying?

GREAT DISADVANTAGE	SLIGHT DISADVANTAGE	MAKES NO DIFFERENCE	SLIGHT ADVANTAGE	GREAT ADVANTAGE
0	0	0	1	15

$\underline{D}=.74$ ($p < .05$)

COMMENTS:

Subject 1: No comment

Subject 2: No comment

Subject 3: No comment

Subject 4: Speech recognition during heads up is probably its most valuable asset.

Subject 5: No comment

Subject 6: No comment

Subject 7: However--there is not a lot of room on today's HUDs for lengthy presentations on the HUD for non-critical information.

Subject 8: No comment

Subject 9: It highly complements the concept of the HUD and doesn't get the pilot back looking into the cockpit to change a radio or whatever.

Subject 10: It allowed me to more fully devote work to flying and enabled me to keep the large transition from inside the cockpit to the HUD down to a minimum.

Subject 11: No comment

Subject 12: No comment

Subject 13: Especially low level/air-to-air.

Subject 14: No comment

Subject 15: No comment

Subject 16: No comment

12. In reference to the vocabulary used for the speech activated MFC, how would you rate the vocabulary selected for each task?

	UNACCEPT- ABLE	BAD	SATISFAC- TORY	GOOD	OPTIMUM	D
UHF CHANGE	0	0	2	8	6	.48*
TACAN CHANGE	0	0	2	7	7	.48*
BAROMETER CHANGE	4	4	3	3	2	.1
IFF NORMAL	0	5	3	4	2	.2
IFF MODE 1 CHANGE	0	2	5	5	4	.28
IFF MODE 3 CHANGE	0	2	6	4	4	.28
FLY TO ENGAGE	0	1	4	6	5	.34*
FLY TO DISENGAGE	0	1	4	7	4	.34*
FLY TO CHANGE	0	1	4	6	5	.34*

WEAPON DROP MODE CHANGE	0	3	2	5	6	.29
WEAPON FUZING CHANNEL	0	3	2	5	6	.29
WEAPON INTERVAL CHANGE	0	2	2	6	6	.35*
WEAPON QUANTITY CHANGE	0	2	3	5	6	.29

* $p < .05$

COMMENTS:

Subject 1: "BAROMETER" should be changed to "ALTIMETER." I had a lot of trouble with the Mode 1 and Mode 3 phrase. The phrase should be changed to Mode (pause) 1 and Mode (pause) 3.

Subject 2: No comment

Subject 3: Altimeter not barometer.

Subject 4: Saying COMM before IFF to changes to normal is confusing. This doesn't seem to follow the other IFF change patterns.

Subject 5: No comment

Subject 6: Vocabulary: Option 1, 2, 3, etc. is not used in fighters, why not use bombs--single, salvo etc. Barometer change is not used, use altimeter change as in ATC preferred terminology.

Subject 7: UHF--I would like to be able to say 2 sixty-two--not 2(pause)6(pause)2(pause)1. IFF--Put on IFF

			Mode I	
			Mode II	
Pilot says	1. IFF	2.	Mode III	3. ENTER!
			Mode IV	
			Mode C	
			Stand-by	
			Normal	
			Ident.	

Weapons--I don't like the use of Option 1, Option 2, etc. because it leads you to believe you have several different stores onboard. You can easily program the visual "stores loads" display on ground. What is important is airborne selection and tasks. My input for example:

Say "Stores"--get you to weapons load list

Say "Bombs"--selects all bombs loaded (as opposed to missiles or perhaps CBUs if also loaded

Say "Pairs"--
"Singles"-- as appropriate
"Ripple"--

Say "Nose/Tail"--as applicable

Interval--selected during preflight or manually changed in flight

Quantity--loaded on ground, can change in flight

Subject 8: No comment

Subject 9: I feel the shorter the word that depicts what is to be changed is best.

Subject 10: Barometer should be changed to altimeter which is the common use word for that instrument. Both drop mode and fuzing change should have a key word instead of the current system of just saying pairs or nose. This would put these modes more in line with the call-up path of the other items, i.e., Option1--Quantity--12.

Subject 11: The IFF squawk could be changed without going to another page!

Subject 12: No comment

Subject 13: Realistic and operational.

Subject 14: UHF should include the word "DECIMAL" or "POINT" and the ability to recognize compound numbers (e.g., UHF two fifty three point eight).

"BAROMETER"--change to "ALTIMETER"

IFF--should use "SQUAWK" where appropriate

For weapon changes would prefer some method other than "OPTION __" which requires consulting a list to determine which option.

IFF NORMAL--should not require the "COMM"

Subject 15: Change FLY TO to something such as "STEERPOINT CHANGE" or "STEERPOINT"

"BAROMETER"--"ALTIMETER"

To change IFF Code--Say "Mode 1" then the code

Subject 16: For IFF Mode 3 changes, common tendency from my MAC experience is just to "SQUAWK" that code. So, for IFF mode changes, I suggest using the terminology: "SQUAWK"--"(Desired Mode)"--"(New Code)".

13. In reference to the control logic (i.e. the steps you used in accomplishing a task) you used for the speech activated MFC, how would you rate the efficiency of the control logic for each task?

	VERY INEFFI-CIENT	MOD-ERATELY INEFFI-CIENT	SATIS-FACTORY	MOD-ERATELY EFFI-CIENT	VERY EFFI-CIENT	D
UHF CHANGE	0	1	3	2	10	.43*
TACAN CHANGE	0	0	3	3	10	.43*
BAROMETER CHANGE	1	0	2	4	9	.41*
IFF NORMAL	2	4	5	0	5	.11
IFF MODE 1 CHANGE	0	0	8	1	7	.4*
IFF MODE 3 CHANGE	0	0	8	1	7	.4*

FLY TO ENGAGE	0	0	3	2	11	.49*
FLY TO DISENGAGE	0	0	3	3	10	.43*
FLY TO CHANGE	0	0	3	2	11	.49*
WEAPON DROP MODE CHANGE	0	0	4	5	7	.4*
WEAPON FUZING CHANGE	0	0	4	4	8	.4*
WEAPON INTERVAL CHANGE	0	1	2	6	7	.41*
WEAPON QUANTITY CHANGE	0	1	2	6	7	.41*

* $p < .05$

COMMENTS:

Subject 1: No comment

Subject 2: No comment

Subject 3: No comment

Subject 4: No comment

Subject 5: IFF Normal instead of COMM, IFF, NORMAL, could eliminate COMM, and just have IFF, NORMAL, or SQK, NORMAL.

Subject 6: No comment

Subject 7: No comment

Subject 8: No comment

Subject 9: Words like barometer, interval, quantity, are too long to really be efficient in a speech recognition system.

Subject 10: No comment

Subject 11: No comment

Subject 12: No comment

Subject 13: No comment

Subject 14: Constraining the control logic to the logic used in the manual system hurt the overall efficiency of speech activated MFC.

For "directed" changes (i.e., those commanded from outside the aircraft, UHF, IFF, ALtimeter setting, etc.) the syntax should be consistent with the pilot's normal read-back syntax.

Subject 15: Keep the IFF on the master page for mode changes and OFF, NORM, LOW, and STBY.

Subject 16: No comment

14. Do you have any other comments that you would like to make concerning this simulation? Any feedback that you provide would be very helpful in improving the design of speech recognizers and their implementation in future cockpits.

COMMENTS:

Subject 1: No comment

Subject 2: No comment

Subject 3: Try a single syllable word in place of "ENTER" (e.g., "SET").

Subject 4: I recommend that you think about the effectiveness of speech recognizers during times of high stress or emergencies. I suspect that when a pilot is in heavy combat or during a serious emergency his speech patterns would change. This would not only make the speech recognizer ineffective, but it could also add to the confusion.

Subject 5: No comment

Subject 6: No comment

Subject 7: No comment

Subject 8: No comment

Subject 9: No comment

Subject 10: To be an effective device the recognizer will have to be able to understand the various ways an individual says a word. For example, when one puts an imprint on the

tape it is in a cool sterile environment. When one gets task loaded, a poor visibility approach or an engagement your speech pattern changes possibly to a clipped shorter version of words. This clipped version is not recognizable in the current system and I had a lot of difficulty with 2 and 8 because of the different ways one can say these two words. With this design problem cleared up I would love to fly an aircraft with this capability to include the heavies which could benefit from it.

Subject 11: The feedback in the HUD may not be as helpful when flying out the window. Perhaps a small dedicated alphanumeric display on the flare panel would be better.

Subject 12: No comment

Subject 13: Speed on input needs to be faster to increase operational (combat) effectiveness. Time compression may not allow the luxury of slow broken speech inputs. This is a very viable and useful concept that should be perfected and implemented.

Subject 14: No comment

Subject 15: No comment

Subject 16: No comment

References

1. Lea, W. A *Critical issues in airborne applications of speech recognition*. Warminster, Pennsylvania: Naval Air Development Center, 1979. (AD-A084 703).

2. Navon, D. and Gopher, D. On the economy of the human-processing system. *Psychological Review*, 1979, *86*, 214-255.

3. Herscher, M. B. Real-time interactive speech technology at threshold technology. In P. M. Curran, R. Breaux, and E. M. Huff (Eds.) *Voice technology for real-time command control systems application*. Proceedings of a symposium, Moffett Field, California: Ames Research Center, December 1977.

4. Reddy, D. R. Speech recognition by machine: a review. *Proceedings of the IEEE*, 1976, *64*, 501-531.

5. Doddington, G. R., and Schalk, T. B. The computer listens: how good are today's speech recognition systems? *IEEE Spectrum*, September 1981, 26-32.

6. Broadbent, D. E. *Decision and stress*. New York: Academic Press, 1971.

7. Allport, D. A., Atonis, B., and Reynolds. On the division of an attention. *Quarterly Journal of Experimental Psychology*, 1972, *24*, 225-235.

8. Eggemeier, F. T. Some current issues in workload assessment. _Proceedings of the Human Factors Society 24th Annual Meeting_, 1980, 669-673.

9. Norman, D. A. and Bobrow, D. G. On data-limited and resource-limited processes, _Cognitive Psychology_, 1975, _7_, 44-64.

10. Wickens, C. D., Mountford, S. J., and Schreiner, W. _Task dependent and individual differences in dual task performance_ (NBDL-M003). New Orleans, Louisiana: Naval Biodynamics Laboratory, October 1980.

11. Kahneman, D. _Attention and Effort_. Englewood Cliffs, New Jersey: Pretince-Hall Inc., 1973.

12. Harris, S. D. _Human performance in concurrent verbal and tracking tasks: a review of literature_ (NAMRL - Special Report 78-2). Pensacola, Florida: Naval Aerospace Medical Research Laboratory, July 1978. (AD-A060 493).

13. Welch, J. R. _Automatic data entry analysis_ (RADC-TR-77-306). Griffiss Air Force Base, New York: Rome Air Development Center, September 1977. (AD-A045 939).

14. Taggart, J. L., and Wolfe, C. D. _Voice recognition as an input modality for the TACCO preflight data insertion task in the P-3C aircraft_. Monterey, California: Naval Post Graduate School, March 1981. (AD-A105 568).

15. Connally, D. W. Voice data entry in air traffic control (FAA-NA-79-20). Atlantic City, New Jersey: Federal Aviation Administration, National Aviation Facilities Experimental Center, August 1979.

16. Harris, S. D., Owens, J. M., and North, R. A. Human performance in time-shared verbal and tracking tasks (NAMRL - 1259). Pensacola, Florida: Naval Aerospace Medical Research Laboratory, April 1979, (AD-A070 275)

17. Harris, S. D., North, R. A., and Owens, J. M. A system for the assessment of human performance in concurrent verbal and manual control tasks. Behavior Research Methods and Instrumentation, 1978, 10, 329-333.

18. Mountford, S. J. and North, R. A. Voice entry for reducing pilot workload. Proceedings of the Human Factors Society 24th Annual Meeting, 1980, 185-189.

19. Skriver, C. P. Vocal and manual response modes: comparison using a time-sharing paradigm (NADS-79127-60). Warminster, Pennsylvania: Naval Air Development Center, January 1979.

20. Armstrong, J. W., and Poock, G. K. Effect of operator mental loading on voice recognition performance (NPS55-81-016). Monterey, California. Naval Post Graduate School, August 1981.

21. Wickens, C. D., Vidulich, M. and Sandry, D. Factors influencing the performance advantage of speech technology. Proceedings of the Human Factors Society 25th Annual Meeting, 1981, 705-709.

22. Poock, G. K. A longitudinal study of computer voice recognition performance and vocabulary size, (NPS55-81-013): Monterey, California. Naval Post Graduate School, June 1981.

23. Helwig, J. T. and Council, K. A. (Eds) SAS User's Guide. Cary, North Carolina: SAS Institute, 1979.

24. Cox, C.M., Krishnaiah, P. R., Lee, J. C., Reising, J. M., and Schuurman, F. J. A Study on finite intersection test for multiple comparisons of means. In P.R. Krishnaiah (Ed.) Multivariate analysis (Vol. V) Amsterdam: North-Holland Publishing Company, 1980.